On CK, PCK and student dropout in the early phase of math (teacher) education at university

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Pascal Kilian

aus Ulm

Tübingen

2018

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	06.07.2018
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Frank Loose
2. Berichterstatter:	Prof. Dr. Augustin Kelava

Danksagung

Ich bedanke mich herzlich bei allen die mich in der Zeit als Doktorand begleitet haben.

Mein besonderer Dank geht an Augustin Kelava und Frank Loose, für die überragende Betreuung und eine immer spannende und witzige Zeit, außerdem an Judith Glaesser für die wertvolle Zusammenarbeit.

Ich bedanke mich auch ganz herzlich bei meiner Familie für die tolle Unterstützung und Aufmunterung.

Für alles möchte ich mich auch ganz besonders bedanken bei Carolin und ihren guten Nerven.

Contents

Detailed contents

Summary

Zusammenfassung

Diese Arbeit behandelt zentrale Themen der Mathematiklehrerbildung. Bereits 1986 und 1987 beschrieb Lee S. Shulman in seinen viel zitierten Artikeln eine mögliche Gliederung des professionellen Wissens einer Lehrkraft. Neben anderen Kategorien hat sich hier vor allem die Unterteilung in Fachwissen, fachdidaktisches Wissen und allgemein pädagogisches Wissen durchgesetzt. Shulman selbst beschreibt die Fachdidaktik als Übergang zwischen der Fachwissenschaft und der Pädagogik mit besondere Relevanz für die Lehrperson. Viele Themen der Lehrerbildung sind in diesem Spannungsfeld zu finden. In der vorliegenden Arbeit wird vor allem die Fachwissenschaft und die Fachdidaktik, inklusive deren Übergang, bzw. Trennung, genauer untersucht. In drei Studien werden hierbei unterschiedliche, in diesem Rahmen relevante Themen behandelt. Zum einen wird eine Unterscheidung von Fachwissen und fachdidaktischem Wissen, sowie die innere Struktur der Fachdidaktik in einer sehr fachnahen Auslegung diskutiert (siehe Kapitel 1). Kapitel 2 behandelt darauf aufbauend verschiedene Voraussetzungen für den Erwerb dieser Wissensformen. Mögliche Gruppenunterschiede, zum Beispiel bezüglich des Studiengangs, werden in die Analysen mit einbezogen. Aufgrund der bekannt wichtigen Rolle des Fachwissens der Lehrkraft für gelungenen Unterricht, widmet sich Kapitel 3 dem Erfolg der Studierenden in der Analysis 1 Vorlesung. Diese Vorlesung stellt als verpflichtende Veranstaltung im ersten Semester, unabhängig des Studiengangs, den Einstieg ins Studium dar. Der Erfolg in der Veranstaltung (gemessen an Abbruchraten) wird durch verschiedene statistische Methoden, unter anderem Überlebenszeitanalysen, untersucht. In diesem Rahmen spielen Prädiktionsmodelle, wie sie zum Beispiel im Bereich des Maschinelle Lernens eingesetzt werden, eine zentrale Rolle.

Um eine Datengrundlage zu schaffen wurde hierfür an der Universität Tübingen ein Projekt durchgeführt, in dessen Rahmen auch Kompetenztests für die Wissensformen entwickelt wurden. Die Ergebnisse werden gegliedert in verschiedene Themenbereiche. Der Zusammenhang zwischen Fachwissen

und fachdidaktischen Wissen konnte in mehreren Analysen bestätigt werden. Dabei kann davon ausgegangen werden, dass fundiertes Fachwissen als notwendige, aber nicht hinreichende Bedingung für fachdidaktisches Wissen gesehen werden kann. Trotz dieses Zusammenhangs war eine Trennung der Wissensformen bereits in der Anfangsphase der universitären Ausbildung möglich. Dies deutet darauf hin, dass die Fachdidaktik, zusätzlich zur Fachwissenschaft, ein mathematisches Verständnis darstellt, welches sich bereits zu Beginn, und im speziellen vor der lehrerspezifischen Ausbildung, nachweisen lässt. Die Ähnlichkeit der Voraussetzungen für den Erwerb der Wissensformen deutet erneut auf einen starken Zusammenhang hin. Ausschlaggebend hierfür sind vor allem Leistungsmaße aus der Schulzeit und das Geschlecht, wohingegen der Studiengang für keine der beiden Testleistungen (Fachwissen und Fachdidaktik) relevant war. Für die Testleistung im Bereich der Fachwissenschaft spielte außerdem die Schulform eine Rolle. Studierende, die ihr Abitur an einem allgemein bildenden Gymnasium erlangten, zeigten hier bessere Testleistungen. Im Gegensatz zu den vorherigen Analysen wurde in Kapitel 3 nicht direkt das mathematische Wissen über einen Kompetenztest gemessen. Stattdessen wurden Leistungsmaße aus der Analysis 1 Veranstaltung verwendet, um so den Erfolg in einer Veranstaltung zu untersuchen, die relevant für den Erwerb des mathematischen Wissen und den Einstieg in das Mathematikstudium im Allgemeinen ist. In den Ergebnissen lässt sich die bereits erwähnte Abhängigkeit zu den Leistungsmaßen der Schule (Abiturnote und Note in der Mathematik Abiturprüfung) wiederfinden. Für die Prädiktionsmodelle musste die Note in der Mathematik Abiturprüfung teilweise durch zusätzliche Leistungstests abgesichert werden. Auch hier war die Schulform relevant für den Erfolg im ersten Semester, aber im Gegensatz zu den vorherigen Analysen trat auch der Studiengang als Prädiktor auf. Die Prädiktionsmodelle zeigen Vorhersagegenauigkeiten auf Testdaten von 75%. Zu beachten ist hierbei, dass keine Kennzahlen der Studierenden nach Beginn der Vorlesung in die Modelle mit aufgenommen wurden. Das bedeutet im Besonderen, dass Variablen bezüglich des Verhaltens während des Semesters, wie Zeitaufwand, Anwesenheit und Gewissenhaftigkeit, für die Erfolgs-, bzw. Abbruchvorhersage nicht berücksichtig wurden. Dies deutet auf eine hohe Relevanz der Eingangsvoraussetzungen hin, welche unabhängig von jeweiligem Verhalten während des Semesters, gute Prädiktoren für den Erfolg zu sein scheinen.

Introduction

When defining a knowledge base for teachers' professional competence a wide range of aspects can be considered. Among other categories Shulman (1986, 1987) named the *content knowledge* (CK), *pedagogical content knowledge* (PCK) and *general pedagogical knowledge* (GPK), which were adopted and accepted in several frameworks and literature (e.g., Baumert & Kunter, 2006; Blömeke, 2005b). More specific for math teachers the categories *mathematics content knowledge* (MCK) and *mathematics pedagogical content knowledge* (MPCK) were established (see Schoenfeld & Kilpatrick, 2008). Here I summarize math teachers' professional competence using those categories, with more details in the respective chapters. In math teacher education the categories can be linked to different phases. The initial phase is dominated by and centered on teachers' MCK in math lectures. During the education the focus shifts towards GPK, for example general teaching related contents and classroom management. This shift can be illustrated as it occurs throughout the phases within MPCK, as a blending of MCK and GPK (Shulman, 1987). Using the division of professional competence in stable and situation-specific teacher cognition (Blömeke, Busse, Kaiser, König, & Suhl, 2016), I illustrate the shift in teacher education comparable to a transition from the former to the latter. MCK and MPCK are assigned to stable cognition. MCK is seen as a prerequisite for MPCK, and MPCK itself as a different kind of mathematical understanding (e.g. Baumert & Kunter, 2011; Shulman, 1986). Due to this relationship the initial phase is not only dominated by MCK but also by the content related part of MPCK.

Within this content related framework students face a problem named 'double discontinuity' by Klein (1908) which concerns both, MCK and MPCK.

This refers to potential problems teacher candidates experience on how mathematical contents are taught in school to how they are taught at university and then transferring their knowledge back to school again as a teacher. For students, there is a gap between those two approaches to mathematics, and they gradually come to believe that they are not going to need the contents they studied at university for their future work as teachers. The first discontinuity concerns the transition from school to university. Most of the students will experience a completely different approach to mathematics at the university, from the one they were familiar with at school. At the beginning of their training at the university students face problems, which seem completely different to what they studied at school. For example, in school the focus lies on solving practical mathematical problems with the necessary techniques, whereas at the university students learn about the abstract framework behind those solutions. Students end up thinking that the topics they studied at school are not relevant any more. Albeit its relevance, this thesis will not concentrate on this part of the problem, but rather concentrate on teacher education at the university and therefore the second part of this discontinuity (see e.g Winsløw & Grønbæk, 2014). It concerns the problem of how math teacher candidates can benefit from the academic approach to mathematics at the university and use it to teach in a way which is suitable for school students. If this second part of the discontinuity problem is not resolved, novice teachers will not be able to benefit from the expertise they developed at the university, but rather resort to their own schooling.

This thesis examines the initial phase of math teacher education at university and thus concerns further investigations of MCK and MPCK. Therefore I use results of former studies, like the general importance of MCK for successful teaching demonstrated by Ball, Lubienski, and Mewborn (2001). Krauss, Neubrand, Blum, and Baumert (2008) reported results about the strong relation of MCK and MPCK, and in Kunter and Baumert (2011) this relation is further specified as MCK being a necessary but insufficient requirement for MPCK. Those results are rather general and participants range from teacher candidates of different school types or different phases of teacher education to teachers in service. Interesting and new ideas were introduced by the *mathematical content knowledge for teaching* (Ball, Thames, & Phelps, 2008; Hill, Ball, & Schilling, 2008; Hill, Rowan, & Ball, 2005; Hill, Schilling, & Ball, 2004). Ball et al. (2008) described the idea of a special kind of mathematical understanding needed for teaching. I explain and refer to this framework and the included ideas throughout the thesis. A broad overview of the research in the field of MPCK can be found in Depaepe, Verschaffel, and Kelchtermans (2013).

This study concentrates on the education of math teacher candidates for the academic track and solely on the initial phase of university. This allows more specific investigations on MCK and MPCK. This framework is relevant for math departments at universities, where in the initial phase the MCK and in most cases the MPCK training is located.

Within this framework the present studies cover three topics which are relevant for teacher education at university. First, I investigate the structure of MPCK as well as its relation to MCK. For the education it is relevant to know if – even in this subject matter dominated phase – facets within MPCK can be identified and thus be trained. Additionally, a statistical separation of a content related MPCK and MCK is relevant for the estimation of the importance of additional MPCK lectures and seminars already at the beginning of education. Those considerations can be summarized in the research question: Is it possible to identify facets of MPCK during the initial phase of university, which is dominated by subject matter, and is it possible to statistically separate these facets from MCK? This topic is covered in chapter 1.

Second, for the planning of lectures the knowledge about the prerequisites students bring from school in both, MCK and MPCK, is worth knowing. Differences between teacher candidates and Bachelor of Science students are in this framework particularly interesting. As already mentioned, skills in MCK and MPCK are important for successful teaching, thus I further investigate the question about the existence of a relation between those two facets of knowledge and, if there is one, how strong it is for novice teachers at university. Because of an expected relationship, I investigate the prerequisites students bring to university for the acquisition of MCK and MPCK in chapter 2.

Third, because of the importance of MCK for teaching (e.g. Ball et al., 2001; Krauss et al., 2008), I analyze the success of teacher candidates in a first semester mathematical lecture, compared to Bachelor of Science students. Those issues include the question, if there are differences in the prerequisites of students at the beginning of university, with special interest on the group of teacher candidates. Motivation for this topic gives the supposed negative selection of teacher candidates, which are assumed to be weaker students than their colleagues (Blömeke, 2005a). Results are compared with more general findings about differences of prerequisites at the beginning of university (Klusmann, Trautwein, Lüdtke, Kunter, & Baumert, 2009). Beside prerequisites high dropout rates in math study programs are investigated (Heublein, 2014), by identifying variables and characteristics, which are most important for accurate dropout predictions. Results can be useful for universities to identify risk groups which might need more assistance or to employ admission restrictions. Because of the importance of MCK a focus lies on the question, if teacher candidates can be identified as a risk group in the subject matter education. An extensive analysis of those questions and student dropouts in general is provided in chapter 3.

Structure of the thesis

In the following section the Maths Teacher Education Study (MatTES) is briefly introduced. This project was created to form the database for the subsequent chapters. More details on the project and the parts that are used in the study are given in the respective chapters. In chapter 1, the study for the first topic – the structure of MPCK – is presented. The chapter represents a stand-alone study and has been submitted to an international journal (Kilian, Glaesser, Loose & Kelava, submitted manuscript, 2017). Chapter 2 concerns prerequisites for the acquisition of MCK and MPCK. The contents of this chapter are submitted in German to a scientific journal (Glaesser, Kilian et al., submitted manuscript, 2018). Chapter 3 presents extensive investigations about students' prerequisites and dropout rates, as well as variable selections for dropout predictions.

All chapters are stand-alone studies and can be read independently. Theoretical background information, used methods and samples as well as detailed discussions are provided within the chapters. In the final conclusion the results of all three chapters are summarized and discussed.

The Maths Teacher Education Study – MatTES

The subsequent studies in this thesis rely on a database gathered at the University of Tübingen. For this purpose the project Maths Teacher Education Study (MatTES) was created, which was conducted the first time in the winter semester 2014/15.

The goal of the project is to monitor math students in the initial phase of university, with focus on competence modeling and dropout rates, especially in math teacher education. The concentration is on the analysis lectures and cohorts are followed from the Analysis 1 lecture in the first semester until the Analysis 4 lecture in the fourth semester. Beside the linear algebra, the mandatory analysis lectures play a crucial role at the beginning of mathematics studies and thus qualified as framework for the research questions. Additional data for validation and comparisons is collected in seminars concerning MPCK. The participation in these seminars is mandatory for teacher candidates.

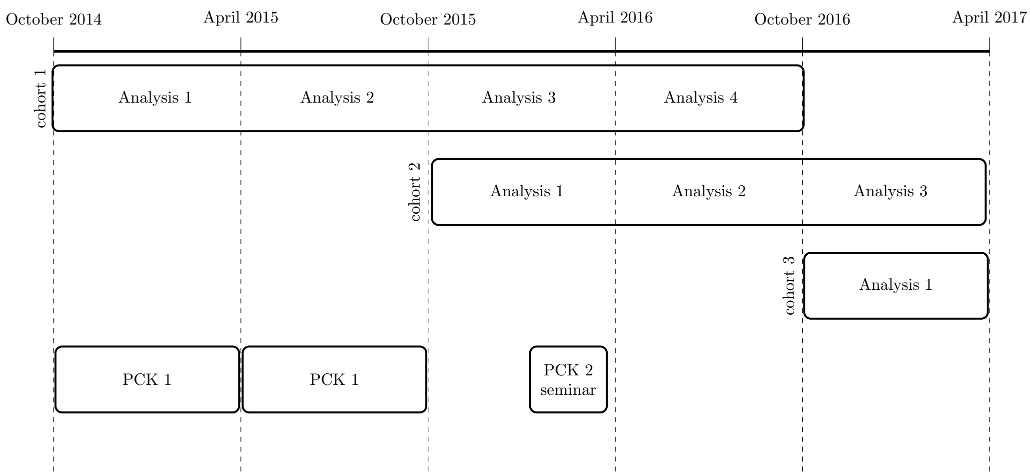

Figure 1: Overview of the MatTES dataset used in the following chapters. Cohort 1 started with the Analysis 1 lecture in winter semester 2014/15 and finished with Analysis 4 in summer semester 2016. The subsequent cohorts two and three started in winter semester 2015/16 and in winter semester 2016/17 respectively. In winter semester 14/15 and summer semester 2015 PCK 1 lectures and in February 2016 a PCK 2 seminar were included.

Instruments and participants

The schedules in the analysis lectures are similar for all cohorts. Additionally to attending the lecture the students participate in tutorial groups, were weekly exercise sheets are discussed. Those exercise sheets have to be solved as homework and are graded. Achievements on those exercises and in the tutorial groups are necessary to gain the admission for the final exam.

In every lecture students complete a questionnaire within the first two weeks. Beside general background information, the questionnaires include tests for both MCK and MPCK. These tests were constantly developed and improved throughout the project. Figure 1 shows an overview of the dataset until the end of winter semester 2016/17. Suitable parts for the respective research questions were drawn from this dataset.

In addition to the tests on MPCK and MCK, the dataset consists of general information on students characteristics like age and sex, performance measures from school like the grade point average (GPA), the math grade in the final exam at school and general information like the school type and the federal state of the school. This information form the prerequisites students bring to university. Information about performance and behavior at the university is collected as described before with student's approval. This includes information about the performance on the weekly exercises, attendance in the weekly tutorial groups and results of final exams and tests.

References

Ball, D. L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (Vol. 4, pp. 433–456). New York, NY: Macmillan.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: what makes it special? *Journal of Teacher Education*, *59*(5), 389–407.

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, *9*(4), 469–520.

Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 29 – 53). Waxmann.

Blömeke, S. (2005a). Das Lehrerbild in Printmedien. *Die Deutsche Schule*, *97*(1), 24–39.

Blömeke, S. (2005b). *Lehrerausbildung, Lehrerhandeln, Schülerleistungen: Perspektiven nationaler und internationaler empirischer Bildungsforschung; Antrittsvorlesung 10. Dezember 2003.* Humboldt-Univ.

Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education*, *56*, 35–46.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher*

Education, *34*, 12–25.

Glaesser, J., Kilian, P., Hische, C., Walz, J., Loose, F., & Kelava, A. (2018). Fachwissenschaftliche und fachdidaktische Kompetenzen bei angehenden Gymnasiallehrer/innen der Mathematik. [Content knowledge and pedagogical content knowledge of trainee teachers for the academic track]. Manuscript submitted for publication.

Heublein, U. (2014). Student dropout from german higher education institutions. *European Journal of Education*, *49*(4), 497-513.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: conceptualizing and measuring teachers' topic-specific knowledge of students [Journal Article]. *Journal for Research in Mathematics Education*, *39*(4), 372–400.

Hill, H. C., Rowan, R., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *41*, 371–406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11–30.

Kilian, P., Glaesser, J., Loose, F., & Kelava, A. (2017). Structure of pedagogical content knowledge in maths teacher education – initial results of the Maths Teacher Education Study (MatTES). Manuscript submitted for publication.

Klein, F. (1908). *Elementarmathematik vom höheren Standpunkte aus. Teil 1: Arithmetik, Algebra, Analysis.* Teubner.

Klusmann, U., Trautwein, U., Lüdtke, O., Kunter, M., & Baumert, J. (2009). Eingangsvoraussetzungen beim Studienbeginn: Werden die Lehramtskandidaten unterschätzt? *Zeitschrift für Pädagogische Psychologie*, *23*(34), 265–278.

Krauss, S., Neubrand, M., Blum, W., & Baumert, J. (2008). The professional knowledge of german secondary mathematics teachers: Investigations in the context of the COACTIV project. In *11th International Congress on Mathematics Education (ICME 11).* Monterrey, Mexico.

Kunter, M., & Baumert, J. (2011). Das COCATIV-Forschungsprogramm zur

Untersuchung professioneller Kompetenz von Lehrkräften - Zusammenfassung und Diskussion. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 345 – 366). Waxmann.

Schoenfeld, A. H., & Kilpatrick, J. (2008). Toward a theory of proficiency in teaching mathematics. In D. Tirosh & T. Wood (Eds.), *International handbook of mathematics teacher education* (Vol. 2, pp. 321–354). Rotherdam, Netherlands: Sense Publishers.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review*, *57*(1), 1–23.

Winsløw, C., & Grønbæk, N. (2014). The ecological Klein's double discontinuity revisited: contemporary challenges for universities preparing teachers to teach calculus. *RDM*, *34/1*, 59–86.

Chapter 1

Structure of pedagogical content knowledge in maths teacher education – Initial results of the Maths Teacher Education Study (MatTES)

In this chapter I discuss the first topic mentioned above. Thereby I focus on content knowledge (CK) and pedagogical content knowledge (PCK) at the beginning of teacher education at university. A key challenge in maths education research is the identification of CK and PCK described by Shulman in 1986. In this chapter, I present a content-related, theoretical framework for mathematical pedagogical content knowledge at the beginning of teacher education. In the present study I use a group of German pre-service teachers – as part of the MatTES dataset (see figure 1) – to investigate both of these dimensions and a further differentiation of pedagogical content knowledge. The empirical results of a bifactor model support a the existence of those different facets. The model is used to illustrate the within-structure of PCK. The model's validity is discussed referring to different types of students and to relevant validity coefficients of scales of other studies. Implications for the

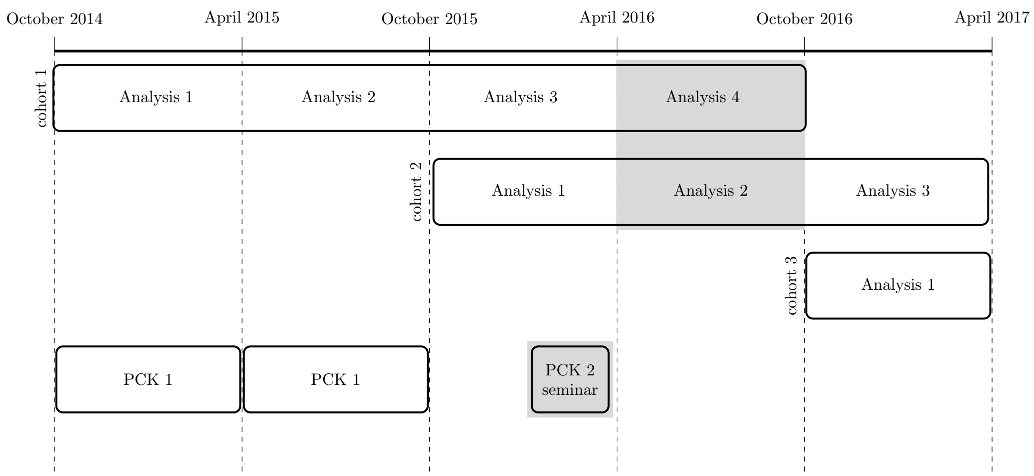

Figure 1.1: Integration of the study *structure of pedagogical content knowledge in maths teacher education* in the design of the Maths Teacher Education Study (MatTES). Highlighted lectures and seminars are part of the analysis.

theoretical foundation of the organization of teacher training are discussed.

This chapter represents investigations in this framework and therefore represents a study on its own, meaning that all abbreviations and the theoretical foundations and integrations are provided. This means that the most important background information is repeated or more specific details are given to provide the isolated readability of this chapter.

The chapter is submitted, with slight changes, to an international journal (Kilian, Glaesser, Loose & Kelava, submitted manuscript, 2017). Authors of the submitted article are: Pascal Kilian; Judtith Glaesser, Frank Loose and Augustin Kelava (in that order). The chapter and the article are written on my responsibility.

1.1 Introduction

In maths teacher education students face a problem termed ‘double discontinuity‘ by Klein (1908). This refers to potential problems experienced at the transition from the way mathematical contents are taught in school to how they are taught at university and vice versa. In this study, the focus is on the second part of this discontinuity (see e.g. Winsløw & Grønbæk, 2014) which concerns the problem of how maths teacher candidates can benefit from the academic approach to mathematics.

Another perspective on this problem is addressed as the 'expert blind spot hypothesis' in Nathan and Petrosino (2003). This hypothesis deals with the problem that advanced knowledge in a content area can lead to notions about learning that are in conflict with students' actual developmental process. Thus this hypothesis describes the gap of mathematical approaches between school and university as described by the double discontinuity problem. Even though novice teachers in maths do not satisfy the definition of experts in the content area, the high level mathematics lectures attended by them at university tend to build this expertise.

This study addresses this problem by investigating the structure of pedagogical content knowledge as something which might be suitable for supplementing the content knowledge education at university. I also examine the lack of differentiation first within the spectrum of pedagogical content knowledge and second within the specific part of the spectrum that can bridge the gap mentioned above.

With this goal I don't see pedagogical content knowledge as only focusing on how to teach mathematics but also as supporting an understanding of mathematical contents that benefits from the high level taught at university and enables novice teachers to bridge the gap between it and the mathematical approach used in school. This potential role it may play is not frequently considered, even though it is not a new notion.

Shulman (1986) explored the question of how college students transform their expertise in the subject matter into a form that is suitable for teaching in school. He suggested a general knowledge base for teachers (Shulman, 1987). Within this, the distinction between *general pedagogical knowledge* (GPK), *content knowledge* (CK) and *pedagogical content knowledge* (PCK) has become accepted in the literature (e.g., Baumert & Kunter, 2006; Blömeke, 2005). The latter is of particular interest because "it represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to the diverse interests and abilities of learners, and presented for instruction" (Shulman, 1987, p. 8). One of the aspects covered by PCK is to comprise the ability of transferring abstract expertise to the sort of knowledge which is relevant for teaching. According to Shulman (1987), PCK may be seen as the connection between content and pedagogy or as the bridge between content knowledge and the practice of teaching. With this point of view, PCK covers a range of areas from content-related elements of teacher knowledge (referred to as the content-related part of PCK) to those elements relating to students and classroom which may be understood as the pedagogical end of the spectrum. Despite its popularity, the wide range of the term – as used by Shulman – lacks definition and empirical investigation (Ball, Thames, & Phelps, 2008). The broad spectrum of PCK indicates a multidimensional construct and it therefore seems appropriate to use a model that takes this dimensionality into account.

The lack of theoretical and empirical grounding of Shulman's conceptualization of PCK is also mentioned by Depaepe, Verschaffel, and Kelchtermans (2013) in their systematic review of research on PCK. Another point of criticism refers to Shulman's static view on teachers' PCK instead of a dynamic view, which treats PCK as inseparable to the act from teaching within a particular context (Depaepe et al., 2013). I share this criticism and support the opinion that PCK in its use is inseparable from teaching situations and a dissociation from other aspects of teachers' knowledge (for example CK) may seem artificial. Nevertheless, like Ball et al. (2008), I think the static point of view is important to identify and isolate different aspects of knowledge. Investigations in this field are necessary to improve and organize teachers' education. Therefore my investigations on aspects and facets of teachers' knowledge are rather static as well.

Lack of differentiation within PCK

This study only focuses on those content-related parts of PCK which may be suitable for contributing to alleviating the second part of the double discontinuity problem. With regard to teachers' knowledge, the importance of subject-matter knowledge itself (e.g. Ball, Lubienski, & Mewborn, 2001) as well as its importance as a prerequisite for PCK (e.g. Kunter & Baumert, 2011) has been noted by various authors. Even though content-related facets are not sufficient to describe the complex structure of teachers' skills and knowledge, those facets are important for subject-matter education at the beginning of university. The concentration on its content-related tasks is a new approach of investigating PCK, as this aspect formed only a small part of the investigation of a general picture in earlier studies (e.g. N. Buchholtz, Kaiser, & Stancel-Piatak, 2011), or it was assigned to CK (e.g. Ball et al., 2008). The structure and the internal differentiation of this aspect of PCK in particular has hardly been studied.

During the last decades, several studies on the relation of CK and PCK were conducted to provide some of the empirical evidence which was considered to be lacking by Baumert and Kunter (2006). Those studies measure CK and PCK separately or their relation to each other among other aspects of the broad professional competence of teachers of mathematics (e.g. COACTIV (Kunter et al., 2011), TEDS-M (Blömeke, Hsieh, Kaiser, & Schmidt, 2014), MT21 (Blömeke, 2011; Blömeke, Kaiser, & Lehmann, 2008) and Blömeke, Busse, Kaiser, König, and Suhl (2016); Krauss, Brunner, et al. (2008)). They are concerned with teachers in service (Löwen, Baumert, Kunter, Krauss, & Brunner, 2011) or students at the end of education (Blömeke & Kaiser, 2014) and focused on contents for lower grades (e.g. Blömeke & Kaiser, 2014; C. Buchholtz et al., 2011; N. Buchholtz et al., 2012).

In those studies, the relationship between CK and PCK was measured using only a single scale for PCK. Facets within PCK were merely formulated for the construction of the test, but have not been empirically verified (e.g. N. Buchholtz et al., 2011). The question about the dimensionality of PCK remains open, but this point is of special interest at the beginning of teacher education because it may be important to know which parts develop together and should be taught together.

Besides other reconceptualizations of teachers' PCK (e.g. Cochran, DeRuiter, & King, 1993; Grossman, 1990; Marks, 1990), Ball et al. (2008) introduced the notion of *mathematical content knowledge for teaching* (MKT) (see also e.g. Hill, Ball, & Schilling, 2008), which covers both CK and PCK. Ball defines MKT as "the mathematical knowledge needed to carry out the work of teaching mathematics" (Ball et al., 2008, p. 395). The framework of MKT reveals interesting insights in teachers' knowledge and is discussed later.

Studies exist with in-service teachers or with student teachers who are close to the end of their studies which aimed to provide a comprehensive picture of teachers' PCK, albeit one which represents PCK as only one construct. However, it may be more appropriate to conceive of PCK as comprising several facets. In the interest of a possible contribution to the improvement of teacher education, it is worth going back one step in order to investigate the structure of PCK at an early stage of teacher training and its development. The question then arises which, specific facets of PCK may already be present at a stage at which the students are still in the process of developing their mathematical foundations. Because those facets might be able to support an understanding of the academic approach to mathematics that will help the students in terms of the transition concerning the mathematical content, the focus has to be on subdivisions of PCK, its different aspects and its relation to CK as they study it in the content-related maths lectures.

1.1.1 Structure of this chapter

First, the next section gives a brief overview of the theoretical frameworks of other studies. Most of those frameworks are intended to provide a general picture of teachers' professional competences.

Secondly, this study, the Maths Teacher Education Study (MatTES) and its framework is introduced. I discuss parallels of MatTES with former studies as well as differences between them. Thirdly, I formulate the research questions. The methods and results sections follow, and I close with a discussion of the findings.

1.1.2 Theoretical framework of former studies

Concepts concerning teachers' professional knowledge

The general framework in the Teacher Education and Development Studies (e.g. TEDS-M, TEDS-shortM and TEDS-LT) splits the professional competence of a teacher into two aspects (e.g. Blömeke et al., 2011; Blömeke, Kaiser, & Lehmann, 2010; N. Buchholtz et al., 2012; Döhrmann, Kaiser, & Blömeke, 2012), an affective-motivational aspect and a cognitive aspect – the professional knowledge. A similar approach was used by the COACTIV group (Professional Competence of Teachers, Cognitively Activating Instruction, and Development of Students' Mathematical Literacy (e.g. Krauss, Neubrand, Blum, & Baumert, 2008; Kunter et al., 2011)). Their aim was to investigate the concepts of professional knowledge and competence of teachers both theoretically and empirically, by bringing together the approaches by different researchers (e.g. Shulman, 1986, 1987; Weinert, 2001) in one model that summarizes several findings and then adapted them specifically for teachers of mathematics (Baumert & Kunter, 2011). Note that both studies divide professional competence into *aspects* of competence, including the professional knowledge (Baumert & Kunter, 2011), which is called cognitive abilities in Döhrmann et al. (2012). Those aspects were divided into different *areas of knowledge*. The focus in this chapter lies on the professional knowledge of a teacher (cognitive abilities) in the context of education students. With (Ball et al., 2008; Hill et al., 2008; Hill, Rowan, & Ball, 2005; Hill, Schilling, & Ball, 2004), a further development of Shulman's conceptualization of PCK is introduced. In contrast to Shulman's notion of PCK, MKT does not arise from purely theoretical considerations, but "resulted from an attempt to refine and empirically validate PCK" (Depaepe et al., 2013, p. 13). Additionally, CK and PCK are not seen as distinct categories, but as part of an overlying category of knowledge, the MKT.

Areas of maths teachers' professional knowledge

In Blömeke et al. (2016), statistical investigations resulted in a model that differentiates between stable and situation-specific teacher cognition. In my view, this differentiation may be used to describe different phases of teacher education. In the early stages – in parallel with the subject-matter training at university – the students primarily acquire stable forms of cognition. Situation specific cognition is more closely related to practice and pre-service in school. The main frameworks of the TEDS and COACTIV groups follow Shulman's suggestion, with professional knowledge separated into CK, PCK and GPK (Blömeke et al., 2016; Bromme, 1992; Shulman, 1987).

In the Teacher Education and Development Study in Mathematics (TEDS-M: Blömeke et al., 2014; Blömeke et al., 2010), which builds on results of TIMSS (Mullis et al., 2007), different components of teacher education were considered. Only one of them deals with mathematics and teaching knowledge (e.g. Tatto et al., 2008).

For the component that includes content-specific knowledge, the framework of Schoenfeld and Kilpatrick (2008) was used. This framework was inspired again by the concepts of Shulman (1987) and distinguishes between *mathematics content knowledge* (MCK) and *mathematics pedagogical content knowledge* (MPCK). Scales for those two parts were developed using item response theory.

Without doubt, CK and PCK are commonly regarded as the main parts of professional knowledge and it is generally agreed on that both include more than the knowledge of contents taught in school. This can also be found in MKT, which consists of facets assigned to either Shulman's CK or PCK (note that the category *curriculum knowledge* is provisionally assigned to PCK, whereas in Shulman (1987) this builds a distinct category). As my focus is on maths specific CK and PCK, I use the abbreviations MCK and MPCK. Again, the aim of this study is to investigate the inner structure of the construct of MPCK.

Facets of MCK and MPCK

In TEDS-M, a differentiation within each of MCK and MPCK was only undertaken for the purposes of item construction (Blömeke & Kaiser, 2014, p. 26). MPCK is subdivided into two facets, with one part focusing on curricular knowledge and planning for mathematics teaching, and another interactive part entitled ‘enacting mathematics for teaching and learning’ (Döhrmann et al., 2012; Tatto et al., 2008). But the subdivisions were then not modelled in the analysis.

The Teacher Education and Development study: Learning to Teach, TEDS-LT (Blömeke et al., 2011) – a national extension of TEDS-M – used a different subdivision of MPCK than in the other TEDS programmes. A distinction is made whether the items related to the underlying subject such as mathematics (subject matter didactics) or to educational science/psychology (education didactics). The short version TEDS-shortM (N. Buchholtz et al., 2012) also uses this approach. But again, all those subdivisions were only introduced during test construction. In the analysis, MPCK is represented by one overall scale since the analysis did not differentiate between the specific aspects.

MPCK in COACTIV consists of three facets. First the importance and the potential of mathematical tasks (*knowledge of mathematical tasks*), second, the awareness of students' misconceptions and comprehension difficulties (*knowledge of students' mathematical thinking*) and third, *knowledge of mathematics-specific instructional strategies*, which includes the ability to give alternative forms of explanation and appropriate methods (Baumert & Kunter, 2011). Diagnostics of students' knowledge is included in the second facet. Note that the latter two components were derived directly from Shulman (1986).

For the empirical examination of the dimensionality of the construct, the authors used confirmatory factor analysis (Krauss et al., 2011; Krauss, Brunner, et al., 2008). In this model, MPCK was measured via sum score indicators of the three facets. However, relying on sum scores is not sufficient to detect the inner structure of MPCK (Little, Cunningham, Shahar, & Widaman, 2002). It only serves as a way of measuring a general score for MPCK.

In these studies, MPCK is broken down into different areas that have to be included in the analysis of a general factor. An analysis of the constructs within MPCK was not intended.

In MKT the focus is not directly on the distinction of different facets within MCK and MPCK, but on identifying facets within the general category of MKT. Those facets are then assigned to the categories of MCK and MPCK, for comparison. In their framework Ball et al. (2008) describe four categories and provisionally add Shulman's *curricular knowledge* (on the side of MPCK) and a category called *horizon knowledge* (on the side of MCK/subject matter knowledge). Both of theses added categories are not part of their investigation. The four main categories are (a) *common content knowledge* (CCK) and (b) *specialized content knowledge* (SCK), which are related to teachers' MCK, and (c) *knowledge of content and students* (KCS) and (d) *knowledge of content and teaching* (KCT), which are related to teachers' MPCK. Ball et al. (2008) define CCK "as the mathematical knowledge and skill used in settings other than teaching. ... In short, they must be able to do the work that they assign their students" (p. 399). This mathematical knowledge is not unique to teaching, therefore the name common knowledge, but it is clarified that CCK can nevertheless be advanced mathematical knowledge not everyone has. SCK represents the new idea within the MKT framework. It describes a special kind of mathematical knowledge unique to teachers. The two main categories within MPCK parallel Shulman's key-components of MPCK (Shulman, 1986), *knowledge of instructional strategies and representations* (cf. KCT) and *knowledge of students' (mis)conceptions* (cf. KCS). A helpful example of the facet interaction is given in Ball et al. (2008) (p. 401) for CCK, SCK and KCS. CCK is needed to recognize a wrong answer, with SCK the nature of an error is recognized, the knowledge about common errors and most likely made errors by students is an example of KCS. In the description of SCK it is noticeable that it might also be seen as the mathematical foundation of the two MPCK facets KCS and KCT. A related criticism refers to the theoretical distinguishability of especially SCK and PCK (Petrou & Goulding, 2011). It is questionable if SCK, as the defined in Ball et al. (2008) (p. 400), is separable from PCK.

1.1.3 Theoretical framework of the present study

As part of MatTES, this study intends to investigate the inner structure of MPCK and how it can support teacher education in Germany. This approach differs from that taken in other studies in that it does not intend to measure correlations of MPCK and MCK in general. Instead, the aim of the study is to investigate dimensions within MPCK and to compare mean scores of different groups differentiated by experience and degree programme, with the aim of validating the test.

The study design of MatTES is different from those of the studies mentioned above, because it is focused on two particular aspects. The first is a concentration on the education of teachers for the academic track (upper secondary education). Pre-service teachers preparing to teach at the academic track are educated separately from those who will teach at other types of school (non-academic track). The type of content studied by the former is far closer to academic mathematics. At most German universities, teacher trainees attend the same subject matter lectures as their fellow students who major in mathematics.

The second aspect concerns the stage of the training. In Germany, teachers' pre-service education takes place in two phases. The first phase is the university-based phase. It involves formal education in MCK, MPCK and GPK as well as internships in school in two subjects. The second phase is the practical induction phase. Here, the students spend most of the week at regular schools where they observe classroom interactions and gradually take on responsibility for lessons. The rest of the week is spent at teacher education institutes where educational studies and subject-related pedagogy are taught.

MatTES investigates students at the beginning of the first phase. It is therefore not necessary for the test to be suitable for students at later stages of their training and for teachers in service, which means that more accurate testing of specific facets is possible. Figure 1.2 presents the area of application of MatTES as well as the blending of MPCK from a teacher/content-related part at the beginning of the training with a pedagogical or student/classroom related part at the end of the training. Figure 1.2 does not include any facets within MPCK, but illustrates how MPCK can parallel the shift of focus in teachers' education. This shift applies also for the inner facets introduced later in this chapter.

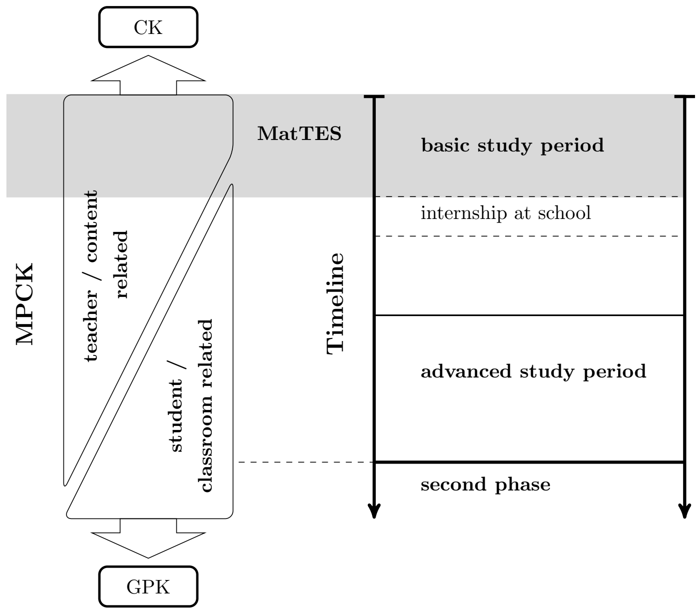

Figure 1.2: Development and blending of the key aspects of mathematical pedagogical content knowledge (MPCK) – from teacher to student related contents – shown next to the timeline of teacher education at university and MatTES' area of application. CK = content knowledge, GPK = general pedagogical knowledge

In the context of MatTES, a definition of MPCK is used which describes the content-related approach to the topic. In this phase a static point of view on the topic (cf. Ball et al., 2008; Shulman, 1986, 1987) is accurate, because the focus lies on factual knowledge gained at lectures and seminars, thus the static point of view is appropriate for the planning of lectures. The dynamic point of view relates to a comprehensive picture of MPCK, including the student and classroom related aspects (see figure 1.2). Here, MPCK is seen as the special kind of subject matter understanding that is necessary to be able to teach mathematics in school (e.g. Baumert & Kunter, 2011). For this kind of understanding, subject matter knowledge is a prerequisite but goes beyond it. MPCK as a different kind of CK, rather than an independent discipline was, reported in Shulman (1986). The notion of a special kind of CK necessary for teachers is well explained as SCK in the MKT framework (Ball et al., 2008). I support this description but assign this notion to the teacher and content related part within MPCK.

In the framework of TEDS-LT, this definition may be compared to the mathematical perspective of MPCK (N. Buchholtz et al., 2011, p. 103, 104). In MatTES, this perspective is not only seen as one part of MPCK, but as the most important aspect at the beginning of training, and for this reason it is the focus of the investigation.

A deep understanding of the mathematical background of the topics taught in school is necessary for teachers, but in addition to that – within MPCK – teachers have to be able to provide different approaches to the topics they teach and understand how students approach maths and what difficulties might appear when introducing mathematical ideas. This includes the ability to assign appropriate tasks to the students (see *tasks* in Baumert & Kunter, 2011), as well as knowledge about students' mathematical thinking (e.g. Baumert & Kunter, 2011; Blömeke et al., 2010). Another special kind of subject matter knowledge within MPCK is the ability to reduce mathematical concepts in a suitable form for school, without changing the original meaning or losing the accuracy required (didactical reduction).

With regard to MCK, this study agrees with the four stages of mathematical understanding in Baumert and Kunter (2011). A profound understanding of the mathematical background of the subject matter taught in school is seen as a prerequisite for MPCK within MCK. In this study I refer to this stage as *school-relevant mathematical content knowledge* (schoolMCK).

This kind of MCK is separate from the highest stage in Baumert and Kunter (2011) (academic knowledge), which is the knowledge taught in subject matter lectures at the university (especially when the lectures are designed for both teacher candidates and students of mathematics as a major subject). This is referred to as *academic mathematical content knowledge* (academicMCK).

I introduce this differentiation not as different facets of MCK, but as a different level of abstraction. At most German universities, the subject matter education for teachers and B.Sc. students is the same at the beginning of their studies. Thus, the subject matter lectures are on a high academical level, which, for example, demands abstract reasoning. This level of abstraction might not be necessary for teachers' MCK. Another reason for the use of different terms is the tasks used to measure MCK. While in the lectures students have to proof their understanding on an abstract level (measured in the exams), in this test the abstraction level was reduced to a teaching level. For simplification, readers are welcome to think of schoolMCK as the CK described in Shulman (1986, 1987) or the CCK in Ball et al. (2008).

Facets within MPCK

In this chapter, MPCK is seen as coming into play in the content-related interaction between the school teacher and his or her students. Therefore, I expect to see two facets of MPCK, which relate to the two directions of this interaction. One facet concerns the direction from the teacher to the students and will be called *instruction*. The other direction concerns the knowledge of cognitive processes of understanding and will be referred to as *diagnostic competence*. Those facets are based on the original key components – knowledge of instructional strategies and representations (*instruction*) and knowledge of students' (mis)conceptions (cf. *diagnostic competence*) – in Shulman (1986). They can be compared to KCT (cf. *instruction*) and KCS (*diagnostic competence*) in the MKT framework. In the MKT framework, I see the components described in SCK as the mathematical foundation of their MPCK facets KCT and KCS. The facets described here include a content-related part and thus include those notions. Figure 1.3 shows the form of this suggested interaction.

The facet of *instruction* refers to the teacher's ability to prepare content and to provide the students with this content. This process includes the integration of the school content in an academic context as well as the extraction and presentation of subject matter (see *tasks* in Baumert and Kunter (2011)) in a form that is suitable for a specific group of students (depending on the

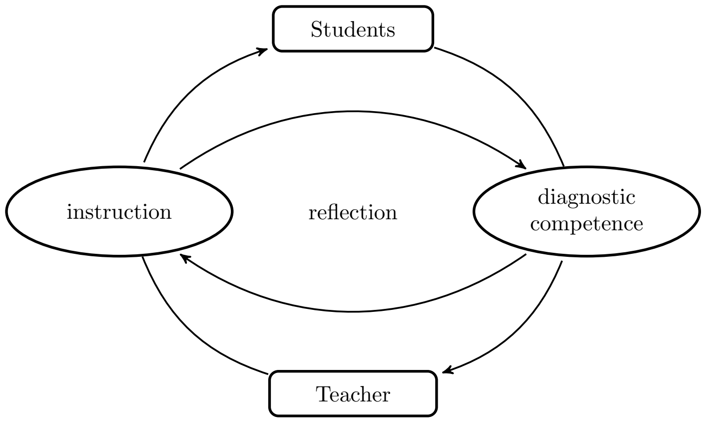

Figure 1.3: Cycle of teacher-student interaction

grade and the previous knowledge of those specific students). Instruction includes knowledge about typical misunderstandings and how they can be avoided by teaching the contents in a particular way or how misunderstandings can be resolved if they arise. Note that I include misunderstandings here, which occur, for example in KCS or Shulman's knowledge of students' (mis)conceptions. I include this notion here, because it is important to understand where problems might occur, for the selection of strategies. In this facets, I also include content-related parts that are described within SCK in the MKT framework. For example, the mathematical knowledge required for the task of "finding an example to make a specific mathematical point" and "modifying tasks to be either easier or harder" (Ball et al., 2008, p. 400) can be seen as the mathematical foundation or the content-related part of the *instruction* facet and thus is included in my framework. Regarding other studies, this facet can be compared to the facets knowledge of mathematical tasks and knowledge of mathematics-specific instructional strategies in the COACTIV framework (Baumert & Kunter, 2011), Shulman's knowledge of instructional strategies and representations and KCT in the MKT framework Ball et al. (2008).

The process which is concerned with the facet *diagnostic competence* includes the identification of underlying sources of mistakes and the evaluation of individual states of knowledge from the responses of the students. The task of identifying the underlying nature of mistakes is described as a component of SCK by Ball et al. (2008). Again, I include this content-related part in the facet within MPCK. In this context, a response can be any kind of feedback the teacher receives from his or her students, for example in oral discussions in the classroom or through test results. In that sense, *diagnostic competence* can be seen as part of the more general diagnostic competence (described for example in Ohle and McElvany (2015)). Close to Shulman's PCK, the general diagnostic competence can be defined as the ability of judging students' performance level correctly as well as the correct estimation of the difficulty of tasks and materials (see e.g. McElvany et al. (2012)). Included in that general framework, *diagnostic competence* qualifies as the part that is concerned solely with the judgement of students' responses to specific contents and tasks. Diagnostic competence is often seen as an important component of teachers' competencies alongside PCK. However, due to the closeness of my notion of *diagnostic competence* to specific contents it is seen as part of PCK. This facet can be compared to Shulman's knowledge of students' (mis)conceptions, KCS (Ball et al., 2008) and knowledge of students' mathematical thinking in COACTIV (Baumert & Kunter, 2011). Figure 1.4 provides a summary of the two facets.

Even though those two facets may be regarded as different aspects of teachers' knowledge, they are not completely separate. The facets are connected through the teacher's process of *reflection* on his or her own actions (Figure 1.3). For example, the evaluation of the state of knowledge in the facet of *diagnostic competence* may affect the preparation of contents within the facet of *instruction*. In my framework the *reflection* does not refer to another facet within MPCK. In order to provide results for the organization of teachers' education I use a static view, but when referring to the practice of teaching, a dynamic interaction of different facets of knowledge is more likely. Therefore the term *reflection* is introduced in figure 1.3 to avoid the notion of distinct facets of knowledge in practice.

In developing the facets of *instruction* and *diagnostic competence*, I drew on the TEDS-M framework (Blömeke et al., 2010), the COACTIV framework (Krauss, Neubrand, et al., 2008; Kunter et al., 2011), the MKT (Ball et al.,

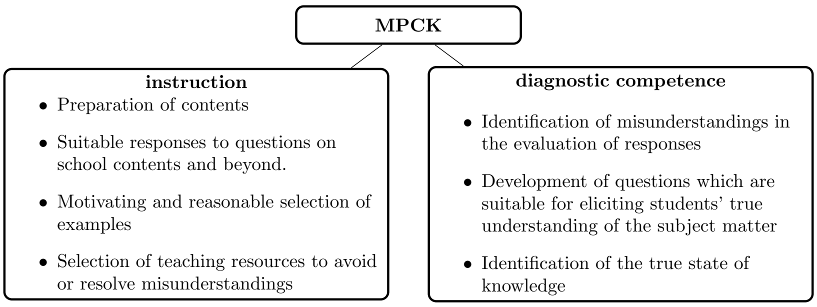

Figure 1.4: Two facets of MPCK

2008; Hill et al., 2008, 2005, 2004) and Shulman. However, compared to the first two studies, my focus is tighter due to the concentration on a particular group of trainee teachers and the differentiation within MPCK. Those facets were tested theoretically in discussions with experts, both on MCK (math lecturers at university) and MPCK (lecturers at teacher education institutes). After the theoretical model validation the facets were tested empirically.

Research Question and Hypothesis

This study investigates MPCK's inner structure and its relation to MCK within trainee teachers who are at the beginning of their university studies. The focus in this phase of teacher education is on subject matter education. In Germany, MPCK lectures – if they are attended at all during this phase – frequently do not match the subject matter contents. As subject matter is often seen as a prerequisite (e.g. Kunter et al., 2011) for MPCK, the question arises whether it is possible to identify facets of MPCK, which exist in addition to MCK, during this phase dominated by subject matter. In order to answer this question I test two models. Model 1 (see Figure 1.6) describes the MPCK facets in addition to the prerequisite of a general MCK facet (here schoolMCK). Model 2 (see Figure 1.7) suggests that schoolMCK alone explains the outcomes of the test.

In this context, these facets may be seen as a specific kind of mathematical understanding that "goes beyond knowledge of subject matter per se to the dimension of subject matter knowledge for teaching" (Shulman, 1986, p. 9).

Ball et al. (2008) mention the importance of evidence for a possible multidimensionality of mathematical knowledge for the organization of teachers' education. "Professional education could be organized to help teachers learn the range of knowledge and skill they need in focused ways" (p. 399). In their paper, they refer to SCK as a dimension of subject matter knowledge beside MPCK. I include this notion in the facets of MPCK, thus identifying my facets in a content-related or subject matter dominated phase can translate to identifying SCK in this phase. This study is deliberately scheduled at the beginning of teachers' education in order to identify the multidimensionality of mathematical knowledge (for teaching) in this phase. If the facets can be identified, teachers' education can be improved, regarding the tasks of MPCK lectures and seminars. While student and classroom related topics within MPCK might play a minor role in this phase, the content-related parts of the facets should be supported parallel to the mathematical lectures for the acquisition of MCK.

A summary of the tasks of MPCK lectures in the first phase of maths teacher education is shown in Figure 1.5. One task is to help students reduce the academic math contents they learn in the lectures to form the schoolMCK facet. This means that an MPCK lecture should support students' transformation of abstract knowledge into knowledge which is useful for teaching in school. The second task is to produce the mathematical understanding necessary for the facets *instruction* and *diagnostic competence*. This understanding involves knowledge on misunderstandings and learning processes.

These facets are not only part of the substance of MPCK and can therefore be drawn on in test construction, but they are actually seen as separate constructs within MPCK in this study.

Hypothesis: Facets of MPCK explain outcomes in addition to MCK in a content-related MPCK test. This means that model 1 suggested here (see Figure 1.6) fits the data and the facets explain variance even for controlled schoolMCK, and model fit improves compared to a one-dimensional model (see Figure 1.7) with schoolMCK as single latent variable.

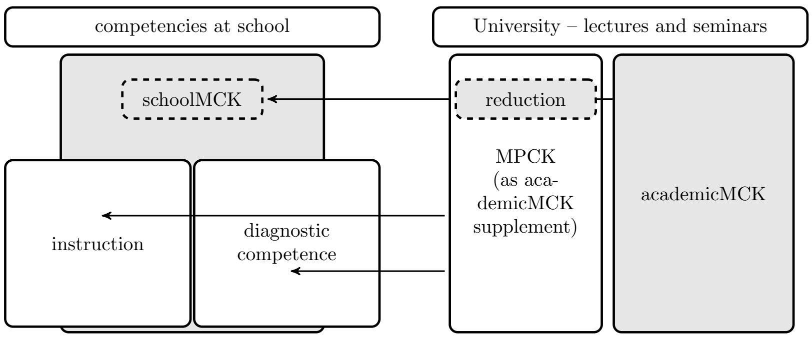

Figure 1.5: Tasks of MPCK (mathematics pedagogical content knowledge) lectures at the early stage in university as support of the subject matter education. MPCK as competence in school composed of the two facets instruction and diagnostic competence, schoolMCK (teachers' school-relevant mathematical content knowledge) as underlying requirement.

1.2 Methods

1.2.1 Study design and sample

The study was conducted within the project MatTES, which started at the end of 2014 at the University of Tübingen. For the research question it was necessary to analyze students with different levels of experience in MCK. I focused solely on MCK in the field of calculus (analysis). A broad calculus training is covered by the lectures Analysis 1 through to Analysis 4, usually attended in the first semester through to the fourth semester. A rough classification would be: One-dimensional integral and differential calculus in Analysis 1, multidimensional differential calculus in Analysis 2, multidimensional integral calculus in Analysis 3 and complex analysis (in one variable) in Analysis 4.

In the maths major course (B.Sc.), it is compulsory to attend all analysis lectures. The state examination students – preservice teachers – (state examination is the German graduate degree for teachers) have to attend Analysis 1, 2 and 4, Analysis 3 is not compulsory for this group of students.

For the teacher candidates, two courses in maths specific didactics are required (Didactics 1 and Didactics 2). Students starting in winter semester

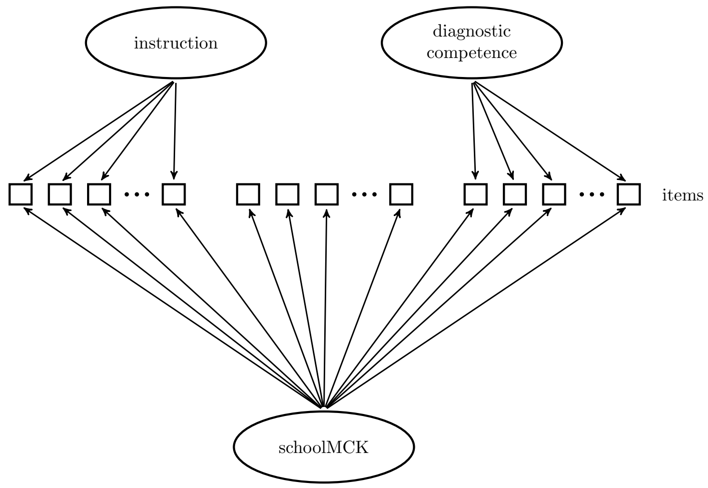

Figure 1.6: Model with schoolMCK (school relevant mathematics content knowledge) and two facets of MPCK (mathematics pedagogical content knowledge) (model 1)

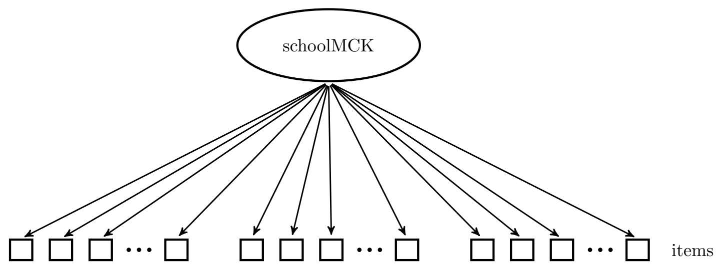

Figure 1.7: One-dimensional model with schoolMCK (school relevant mathematics content knowledge) as single latent variable (model 2)

2014/15 were free to choose at which point during their studies they attended a programme of didactics lectures. Students starting in winter semester 2015/16 were scheduled to take didactics in their second year (third semester), at which point they had already acquired some basic knowledge of academic maths. Therefore, the participants of the Analysis 2 lecture course had not yet attended a didactics lecture course at the start of our study. At the end of winter semester 2015/16, the Didactics 2 course was taught as a one-week intensive seminar.

The empirical study was conducted drawing on participants of two lecture courses and one seminar course. For the main sample, I analyzed students of the two lecture courses Analysis 2 and Analysis 4 (compulsory for teacher candidates) in the summer semester of 2016. These lectures are attended both by students taking either maths or physics as their major subject (in the following we will refer to those two groups together as "B.Sc.") and by teacher candidates (referred to as "Teacher"). Nevertheless this is a homogeneous sample regarding the study programmes because their maths components do not differ at this stage and it is a homogeneous sample within the preservice teachers because until this stage the students mostly focus on the maths education with hardly any additional education in didactics. In addition, students attending the seminar for maths specific didactics (Didactics 2) were included (referred to as "Tsem"). The participants were teacher candidates in the fourth semester and above. Some of them had already completed an internship at school.[1] The sample contained 256 students, 112 of which were recruited from the Analysis 2 lecture and 116 from Analysis 4. The sample from Didactics 2 contained 28 students. In addition to the programmes of study described here (Teacher and B.Sc.), a few students from other programmes attended these lectures which is why the numbers for Teacher and B.Sc. do not add up to the total. Table 3.3 shows descriptive statistics for the samples as well as numbers of teachers in the groups. Covariates like age and school grades were self-reported via a questionnaire.

[1]Even though this does not seem homogeneous all the analysis was executed and the structure was tested without the inclusion of the seminar resulting in no significant differences.

Table 1.1: Descriptive data for the samples and subsamples

event	"Analysis 2"			"Analysis 4"			"Didactics 2"	total
subsample	total	Teacher	B.Sc.	total	Teacher	B.Sc.	Tsem	sample
	(n=112)	(n=63)	(n=31)	(n=116)	(n=50)	(n=59)	(n=28)	(n=256)
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Age[a]	21.37 (4.65)	20.90 (1.64)	22.32 (8.0)	22.29 (2.34)	22.90 (1.92)	21.86 (2.59)	24.12 (1.35)	22.12 (3.53)
	(n=100)	(n=63)	(n=31)	(n=111)	(n=50)	(n=58)	(n=28)	(n=239)
High school GPA[b]	1.82 (0.55)	1.86 (0.55)	1.73 (0.56)	1.87 (0.56)	1.81 (0.56)	1.92 (0.56)	1.89 (0.54)	1.85 (0.55)
	(n=98)	(n=63)	(n=29)	(n=112)	(n=50)	(n=58)	(n=28)	(n=238)
High school maths score[c]	12.34 (2.41)	12.19 (2.35)	12.90 (2.21)	12.64 (2.53)	12.53 (2.02)	12.91 (2.44)	12.20 (2.06)	12.47 (2.43)
	(n=97)	(n=62)	(n=29)	(n=108)	(n=49)	(n=55)	(n=25)	(n=230)
	%	%	%	%	%	%	%	%
Gender (male)	48	38.1	64.52	59.29	36	76.27	32.14	51.15
	(n=100)	(n=63)	(n=31)	(n=113)	(n=50)	(n=59)	(n=28)	(n=241)

Note. [a] rescaled measure (2016 – (year of birth)). [b] German grade point average (GPA), scores range from 1 to 6 with 1 as the best score. [c] maths score of the final secondary school examination. Scores range from 1 to 15 with 15 as the maximum score.

1.2.2 Instrument

The competence test employed here was developed for trainee teachers at the beginning of their second year at university and above. Basic knowledge of university maths is required. The test is intended to measure knowledge in MPCK that is close to purely mathematical understanding. It is not intended to measure general MPCK skills in teaching and classroom management because these skills are expected to be developed during later stages of teacher education. Because of this and the general area of application of MatTES it was necessary to develop a new instrument. I considered employing instruments which had been validated in other studies (e.g., TEDS-M, TEDS-sM). However, these would not have been suitable for my purposes because the target populations differ. The participants' state of knowledge precluded the use of technical terms in the field of MPCK. Nevertheless, the development of the items was drew on those existing instruments and the best subset was selected.

Different experts were involved in the development of the test. On the mathematics side, the group consisted of lecturers involved in the Analysis lectures. For the MPCK content, the research group worked together with colleagues at the teacher education institute Tübingen, which is responsible for the second phase of maths teacher education and which provides the MPCK lectures and seminars at the university.

In the interest of economy, a fairly short test was constructed containing

19 items, all of which are in multiple-choice format. More items – including existing items – were tested during the development phase. Due to limited testing time the resulting instrument contains a selection of the best items. The study was conducted in close connection with the Analysis lectures, therefore the mathematical background for the items is the content of the Analysis introduction lecture programme, taken in the first semester. Example tasks are presented in the appendix A.

Instruction sub-scale

The instruction sub-scale of the test consists of 5 items, which were also identified empirically. In this facet, the test-takers were asked to consider representations of mathematical theories that are suitable for students in school. Reduction of abstract theories as well as knowledge about the learning process and the individual knowledge state of the students are important. In developing the instrument to measure this kind of knowledge I asked questions, for example, about suitable representations of mathematical theories in school.

Diagnostic competence sub-scale

The 5 diagnostic competence items cover the classification of students' responses into taught contents, identification of possible sources of errors, and interpretation of learning processes. These items, again, were also identified empirically. For instance, wrong student responses are presented, and the task is to identify the type or source of the underlying error or misunderstanding. This sub-scale describes a mental representation of dysfunctional cognition which then serves as a starting point for further instruction. For example, test-takers are asked to identify the error, underlying a wrong student response by deciding which of the tasks presented requires the same mathematical understanding as the original one. For this task, the same mistake that led to the original wrong response may occur again and it can be used for practice.

schoolMCK sub-scale

9 items were developed specifically to measure schoolMCK. These cover content knowledge within the field of analysis in a form that is relevant in school. This dimension is seen as a prerequisite for teaching mathematics and for the MPCK facets. The overall rationale behind these items is, that we wished to capture the way participants coped with the problem of balancing the reduction of mathematical contents to a form suitable for teaching in schools, without losing accuracy. This can be addressed for instance by rating unconventional student solutions. For these tasks, the test-takers have to be confident enough in their mathematical knowledge to identify correct answers, even if these are written in unconventional language or hidden in unconventional thoughts. They also have to be able to isolate main ideas of mathematical theories and objects to discuss them with the students, without going into detail of the underlying academic theory.

Validation items

A collection of TEDS-sM items (N. Buchholtz et al., 2012) was used as comparison items. Due to time restrictions, it was not possible to use the full TEDS-sM scale. Nine items within three different tasks were chosen, all from the cognitive dimension *evaluate and create*. The tasks are DS29 and DBJ4 from the topic of *subject matter didactics* and SUG2 from *education didactics*. This choice was made in order to achieve a good fit with the topics of our lectures. Minor language changes had to be made to ensure that all technical terms were known to our students. Technical terms from the field of didactics in particular are unfamiliar to the students at this early stage of their training.

1.2.3 Procedure

The students voluntarily filled in a questionnaire in the second week of the summer semester 2016. The survey was conducted in the tutorials accompanying the lectures which are attended by about 15 to 20 students each.

The questionnaire contained the test and questions relating to background information. The test time was around 35 min altogether.

1.2.4 Scaling and data analysis

In the analysis, I used structural equation modeling (SEM) (see e.g. Bollen, 1989), to fit latent models. First, the bifactor model 1 (see e.g. Reise, 2012) (see Figure 1.6), then the alternative model 2 (see Figure 1.7) was fitted. The analyses were conducted using the R-package lavaan (R Core Team, 2015; Rosseel, 2012).

Assessment of model fit

I tested the fit of the bifactor model 1 and compared it to the unidimensional model 2 to investigate the hypothesis that the two facets of MPCK can explain additional variance in the data. I used the Tucker-Lewis index (TLI), the comparative fit index (CFI) and the root mean square error of approximation (RMSEA) to evaluate goodness of fit (see e.g. Marsh, Hau, & Wen, 2004). Values of TLI and CFI greater than .90 and .95 are usually taken to indicate acceptable and excellent fits to the data, respectively. RMSEA values smaller than .60 would show a reasonable fit. In addition, we report the χ^2 test statistic, the ratio of the χ^2 deviance and the degrees of freedom. χ^2/df estimates < 2 are regarded as very good fit.

To compare the nested models, I followed the suggestions of Chen (2007) that a model should be favoured, if incremental fit indices, such as the CFI, increases by more than .015, compared to the more parsimonious model. In addition, I executed a χ^2- difference test. Although measurement invariance (Meredith, 1993) should be ensured to conduct comparisons between the subgroups (semesters), this was not possible with the present data due to the small numbers of participants in the subgroups.

Comparison of means

Means on the latent variables were calculated for different majors, different cohorts and gender. I expected to see no differences for gender and no differences between the subgroups of the second semester cohort. At this stage, none of the students had attended didactic lectures and the teacher candidates have not yet had any practical experience. Differences may be expected to occur within the fourth semester cohort. Here, the teacher candidates attending the Analysis 4 lecture may have attended a basic didactics course. The students taking the didactics seminar are more experienced in didactics as some of them have already completed the practical phase at school. Therefore, I expect to see differences between B.Sc. students with majors in maths or physics and teacher candidates who are either taking Analysis 4 or the didactics seminar.

Validity coefficients

The predicted values for the latent variables are also used to compute correlations with other scales and school grades. For the scaling of the TEDS-sM benchmark items a one-dimensional model was fitted using lavaan (Rosseel, 2012) and the person parameters were predicted.

1.3 Results

In this section I present the results for our two models. Model 1 (Figure 1.6) is a bifactor model (see e.g. Reise, 2012). The alternative model 2 (Figure 1.7) is a unidimensional g-factor model. I calculated these models to examine whether the instruction and diagnostic competence sub-scales explain additional variance in the data compared to the g-factor model. Therefore, results of the comparison are also presented. As measure for the reliability of the factors schoolMCK, instruction and diagnostic competence, Cronbach's a was used. The values (.8, .72, and .7) were good.

1.3.1 Comparison of two alternative models

The bifactor model 1 (see Figure 1.6) has a good fit to the data, with respect to the relative fit indices (CFI = .935, TLI = .921), absolute fit indices (RMSEA = .023) as well as $\chi^2/df = 1.133$. Table 1.2 presents fit results for both models. The fit of the alternative model 2 (see Figure 1.7) is inferior to that of model 1 (see Table 1.2). The relative fit indices (CFI = .799, TLI = .771) show a substantively worse fit. Absolute fit indices for model 2 (RMSEA = .039) and χ^2/df = 1.39 also indicate that model 1 has a better fit. The CFI difference is ΔCFI = 0.136 (Robust) which favors model 1 as it is greater than .015 (Chen, 2007). The results of the scaled χ^2 Differences Test (see Rosseel, 2012; Satorra, 2000) are presented in Table 1.3. The test is highly significant for model 1.

Table 1.2: Model fit statistics for both models

	fit indices			χ^2 tests		
Model	CFI	TLI	RMSEA	χ^2	*df*	χ^2/df
bifactor model 1	.935	.921	.023	158.55	140	1.133
g-factor model 2	.799	.771	.039	207.766	150	1.39

Note. CFI=comparative fit index, TLI=Tucker-Lewis index, RMSEA=root mean square error of approximation

The wording of some items was similar which may have led to a lack of independence between them. For this reason, in both models two residual covariances within the diagnostic competence facet had to be calculated. Each residual covariance occurs between two items within one task.

Table 1.3: Scaled χ^2 Differences Test

	df	χ^2	$\Delta\chi^2$	Δdf	$\Pr(>\chi^2)$
model 1	140	132.29			
model 2	150	191.81	44.6	7.9908	$4.344 \cdot 10^{-07}$

1.3.2 Comparison of means

Comparisons of latent dimension means for different subgroups were tested. The results are presented in Table 1.4 including Cohen's d for the effect size (Cohen, 1988) and results of one-tailed t-tests.[2] As expected, I found no

[2] Possible gender differences were examined, but none of the results were significant.

significant differences between different majors in the second semester. It is worth noting in particular that there were no differences between teacher candidates and students with maths as their major subject at this early stage in their performance on the subject matter part, the schoolMCK sub-scale. On the other hand, for the fourth semester students, significant differences were found between the students with maths or physics as their major subjects and the more experienced teacher candidates of the didactics seminar. The teacher candidates performed significantly worse in the area of school-relevant content knowledge (t-value=2.0, df=52.76, d=0.46), but significantly better in the instruction sub-scale (t-value=-2.59, df=75.28, d=0.52). On the instruction sub-scale, the teacher candidates in the didactics seminar also outperformed the less experienced teacher candidates in the Analysis 4 lecture (t-value=-1.83, df=72.86, d=0.40).

Table 1.4: Comparison of means table with t statistics

		second semester			
	Teacher (n=63)	B.Sc. (n=31)			
	M (SD)	M (SD)	t-value (df)	p-value	d
sMCK	-0.05 (0.66)	0.03 (0.48)	-0.71 (79.4)	0.24	0.13
I	-0.05 (0.55)	-0.11 (0.58)	0.50 (56.63)	0.31	-0.11
D	-0.13 (0.63)	-0.18 (0.61)	0.34 (60.80)	0.37	-0.08
		fourth semester			
	Teacher (n=50)	B.Sc. (n=59)			
	M (SD)	M (SD)	t-value (df)	p-value	d
sMCK	-0.09 (0.81)	0.03 (0.71)	-0.81 (98.51)	0.21	1.16
I	-0.03 (0.56)	-0.10 (0.57)	0.66 (104.78)	0.25	-0.12
D	-0.02 (0.55)	-0.06 (0.55)	0.41 (104.28)	0.34	-0.07
	B.Sc. (n=59)	Tsem (n=28)			
	M (SD)	M (SD)	t-value (df)	p-value	d
sMCK	**0.03 (0.71)**	**-0.30 (0.72)**	**2.0 (52.76)**	**0.025**	**-0.46**
I	**-0.10 (0.57)**	**0.17 (0.38)**	**-2.59 (75.28)**	**0.01**	**0.52**
D	-0.06 (0.55)	0.04 (0.65)	-0.72 (46.13)	0.24	0.17
	Teacher (A4) (n=50)	Tsem (n=28)			
	M (SD)	M (SD)	t-value (df)	p-value	d
sMCK	-0.09 (0.81)	-0.30 (0.72)	1.18 (61.89)	0.12	-0.27
I	**-0.03 (0.56)**	**0.17 (0.38)**	**-1.83 (72.86)**	**0.04**	**0.40**
D	-0.02 (0.55)	0.04 (0.65)	-0.60 (48.6)	0.35	0.10

Note. sMCK=schoolMCK, I=instruction, D=diagnostic competence, Teacher=teacher candidates, B.Sc.=major in maths or physics, Tsem=participants of the didactics seminar Significant differences are highlighted in bold. Differences occur within the fourth semester groups. At this stage the curricula start to differ (some teacher candidates attended lectures in MPCK).

1.3.3 Validity coefficients

Correlation coefficients were calculated between the latent variables of model 1 (sMCKscore, Iscore, Dscore), the scaled TEDS-shortM items (TEDSscore) and the individual's GPA (German grade point average) and maths score (of the final secondary school examination). The results are presented in Table 1.5. The similarity of the coefficients of sMCKscore (.36 for the maths score and .35 for the GPA) and TEDSscore (.27 for the maths score and

Table 1.5: Table of correlations

	TEDSscore	sMCKscore	Iscore	Dscore	maths score	GPA[a]
TEDSscore	1.00					
sMCKscore	0.37**	1.00				
Iscore	-0.06	0.18**	1.00			
Dscore	0.15*	0.12	-0.10	1.00		
maths score	0.27**	0.36**	0.03	0.05	1.00	
GPA[a]	0.26**	0.35**	0.03	0.04	0.63**	1.00

Note. [a] inverted, * p<.05 (two tailed), **p<.01 (two-tailed)

.26 for the GPA) with the school variables and their own correlation of .37 indicate the measurement of similar constructs.

1.4 Discussion

In MatTES, MPCK was conceptualized with two inner facets – instruction and diagnostic competence – which characterize teachers' interactions with students. The starting point was a content-dominated approach to MPCK at the beginning of teacher education. At this stage of teacher education the emphasis is on the development of MCK. For this reason, such a content-related approach is more appropriate than more general conceptualizations of MPCK, which might also include general pedagogical aspects. The approach employed here classifies MPCK as a kind of mathematical understanding that a teacher needs for interacting with students and it is seen as something different from the mathematical understanding that students gain when attending lectures in pure mathematics at university. This approach can be compared to the notion of SCK (e.g. Ball et al., 2008), which refers to a mathematical understanding unique for a teacher. It involves an understanding that provides the ability to transfer abstract knowledge in a form suitable for teaching in school.

Validity

The validity of the instrument was ensured in various ways. First, experts from the teacher education institute in Tübingen and lecturers involved in the Analysis lectures worked together during test construction to ensure content validity. Secondly, the test results show substantial correlations with the validity items employed in TEDS-sM (N. Buchholtz et al., 2012) and with school grades such as GPA and the final high school maths score. Note that low correlations of Dscore and Iscore (diagnostic competence and instruction) with other scales are due to the fact that I controlled for schoolMCK and those scores can be seen as any effect remaining after controlling for schoolMCK. Thirdly, the mean differences on the scales conform to expectations based on students' study programmes. At the very beginning of the training there are no differences in competencies. Later in the training, differences in performance on the schoolMCK scale were found between students depending on their programme of study – students who major in mathematics or physics perform better than teacher candidates of the didactics seminar – and in the instruction scale – the more experienced teacher candidates who attended the didactics seminar perform better than students attending the Analysis lecture.

Model results

The identification of MPCK facets beside strictly mathematical knowledge within this content-related point of view was carried out by comparing two models – a bifactor model (model 1) (Reise, 2012) with the MPCK facets added to a general dimension of school-relevant mathematical content knowledge (schoolMCK) and a unidimensional model (model 2) without the additional MPCK facets (see Figure 1.6 and Figure 1.7). The results of the model fit analysis favored the bifactor model (model 1), which supports the existence of two MPCK facets in addition to the general schoolMCK facet. The latter is seen as a prerequisite for MPCK. Although the identification of MPCK beside MCK had already been analyzed in previous studies (see e.g. N. Buchholtz et al., 2011), the results are remarkable because the identification of the different MPCK facets and MCK was undertaken in a highly content-related framework. Within this framework the different constructs can be seen as different kinds of mathematical understanding. Compared to these former studies (see e.g. Blömeke et al., 2011; Blömeke et al., 2010; Kunter et al., 2011) my results show not only a general separation of MPCK and MCK for mathematics teachers, but also a separation of MPCK and MCK as kinds of mathematical understanding and the different development of these facets in the initial stages of the training depending on the study programme. By separation, in this context, I don't mean a theoretical separation, in the sense of independent dimension, but a statistical identification of distinct facets. This shows that MPCK and MCK can be identified separately in the sense of Shulman's PCK as "subject matter knowledge for teaching" (Shulman, 1986, p. 9). This goes beyond the separation of MCK and a comprehensive MPCK dimension – including not only content-related parts but also the general pedagogical point of view – at a later point in the training as employed by the TEDS-group (e.g. Blömeke et al., 2011; Blömeke et al., 2010; N. Buchholtz et al., 2012) and COACTIV (Krauss, Neubrand, et al., 2008; Kunter et al., 2011). In their context a separation seems more obvious because they include aspects of MPCK which differ greatly from MCK. In my context the identification of those facets can rather be compared to the identification or statistical separation of SCK and MCK (in the MKT framework (e.g Ball et al., 2008).

For the investigations in this content-related context, I deliberately scheduled the study at this early stage of the training for two reasons. The first is that at that stage, the focus is on developing subject matter knowledge. The second is that study programmes for all students in mathematics are similar at that stage, and in particular there are no lectures and seminars in the MPCK context. The identification of the facets here can then be seen as evidence for the importance of supplemental lectures addressing this kind of mathematical understanding already at this early stage.

Means and development

The results show no differences between teacher candidates and other students at the beginning of the training. Neither in the MCK nor in the MPCK facets did major differences arise. This was expected due to the homogeneous sample and to their having an identical curriculum in the second semester. In the fourth semester, the curricula of the teacher candidates and the other students are different. While students who major in maths concentrate mostly on mathematical lectures, the teacher candidates attend additional lectures and seminars and some of the teachers' sample in the fourth semester had already completed the practical phase in school. The results reflect this specialization in differences of the latent variable means. The more experienced teacher candidates of the didactics seminar outperformed teacher candidates and the other students in the Analysis 4 lecture on the instruction scale. This could have been caused by the experience those students gained in the practical phase at school, the expert monitoring included in this practical phase, as well as by the didactic seminars and lectures. By contrast, the students of the didactics seminar did less well on the MCK facet – schoolMCK – than the students in Analysis who major in mathematics or physics. This is not surprising in view of the differing curricula in later phases of the training. These results indicate a differentiation of competence during the training from a common starting point. Depending on the study programme, the development of competence differs. While the programmes of the major in mathematics and physics focus on the subject matter, the focus in the study programmes for teacher candidates shifts to a parallel development of knowledge in both MCK and MPCK.

The results may be seen as evidence that it is possible to separate different kinds of mathematical understanding – referred to as MPCK and MCK – as early as the beginning of the training. This result also supports the notion of a mathematical knowledge and skill unique to teaching (described as SCK by Ball et al. (2008)) which should be supported parallel to the subject matter education. Even though I include this knowledge in MPCK facets, the result could be translated to SCK in the MKT framework, were it is located on the side of MCK. Furthermore, MPCK can be separated into facets at that stage, which is interesting from a theoretical point of view with respect to the emergence and structure of competencies. This is important because it can help in the planning of lectures and seminars at that phase. In addition to MCK, MPCK is important for teachers and according to the results of this chapter – from a content-related point of view – it starts to develop together with MCK from the beginning of the training. The separation, however, shows that it is not one simple unidimensional construct and thus should be supported in addition to the mathematics lectures early in the training.

Limitations and outlook

The study and the formulation of the problem arise from, and are based on, teacher education in Germany. Thus, the generalization to other countries might be limited. The version of the test applied here was a very first step in measuring MPCK from a content-related point of view. Now that the results have confirmed the desired possibility of separating mathematical knowledge in that framework, the test has to be expanded. In addition, testing time was limited in the study, so the test was rather short. In the test presented here, only multiple choice items were used. It would be desirable to include open response items for the MPCK facets. The intention is to develop a more detailed version of the instrument with the support of a broad group of experts. For further (content) validity examinations the improved test will be conducted with students of alternative study programmes, like chemistry. As a benchmark examination, the improved instrument will then be applied to a) students in the second phase of teacher education and b) teachers in service. Additionally, for the quantification of criterion validity, it should be examined how school students' competencies are affected by teachers' scores on the facets diagnostic competence and instruction competence. This could not be examined in this first study, but will be part of further investigations.

References

Ball, D. L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (Vol. 4, pp. 433–456). New York, NY: Macmillan.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: what makes it special? *Journal of Teacher Education*, *59*(5), 389–407.

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, *9*(4), 469–520.

Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 29 – 53). Waxmann.

Blömeke, S. (2005). *Lehrerausbildung, Lehrerhandeln, Schülerleistungen: Perspektiven nationaler und internationaler empirischer Bildungsforschung; Antrittsvorlesung 10. Dezember 2003.* Humboldt-Univ.

Blömeke, S. (2011). Teacher Education and Development Study: Learning to Teach (TEDS-LT) - Erfassung von Lehrerkompetenzen in gering strukturierten Domänen. In S. Blömeke et al. (Eds.), *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT* (pp. 7 – 24). Münster: Waxmann.

Blömeke, S., et al. (Eds.). (2011). *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT.* Münster: Waxmann.

Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education*, *56*, 35–46.

Blömeke, S., Hsieh, F.-J., Kaiser, G., & Schmidt, W. H. (Eds.). (2014). *International perspectives on teacher knowledge, beliefs and opportunities to learn: TEDS-M results.* Springer.

Blömeke, S., & Kaiser, G. (2014). Theoretical framework, study design and main results of TEDS-M. In S. Blömeke, F.-J. Hsieh, G. Kaiser, & W. H. Schmidt (Eds.), *International perspectives on teacher knowledge, beliefs and opportunities to learn: TEDS-M results.* Springer.

Blömeke, S., Kaiser, G., & Lehmann, R. (Eds.). (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegeneiten deutscher Mathematikstudierender undreferendare; Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung.* Waxmann.

Blömeke, S., Kaiser, G., & Lehmann, R. (Eds.). (2010). *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich.* Münster: Waxmann.

Bollen, K. A. (1989). *Structural equations with latent variables.* Wiley.

Bromme, R. (1992). *Der Lehrer als Experte. Zur Psychologie des professionellen Wissens.* Bern: Huber.

Buchholtz, C., Doll, J., Stancel-Piatak, A., Blömeke, S., Lehmann, R., & Schwippert, K. (2011). Anlage und Durchführung der Studie TEDS-LT. In S. Blömeke et al. (Eds.), *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT* (pp. 25 – 46). Münster: Waxmann.

Buchholtz, N., Kaiser, G., & Stancel-Piatak, A. (2011). Professionelles Wissen von Studierenden des Lehramts Mahtematik. In S. Blömeke et al. (Eds.), *Kompetenzen von Lehramtsstudierendenin gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT* (pp. 101–133). Waxmann.

Buchholtz, N., Scheiner, T., Döhrmann, M., Suhl, U., Kaiser, G., & Blömeke, S. (2012). *TEDS-shortM: Kurzfassung der mathematischen und* *mathematikdidaktischen Testinstrumente aus TEDS-M, TEDS-LT und TEDS-Telekom.* Hamburg: Universität.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504.

Cochran, K. F., DeRuiter, J. A., & King, R. A. (1993). Pedagogical content knowing: An integrative model for teacher preparation. *Journal of Teacher Education*, *44*(4), 263–272.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, *34*, 12–25.

Döhrmann, M., Kaiser, G., & Blömeke, S. (2012). The conceptualisation of mathematics competencies in the international teacher education study TEDS-M. *ZDM*, *44*(3), 325–340.

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education.* New York: Teachers College Press.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: conceptualizing and measuring teachers' topic-specific knowledge of students [Journal Article]. *Journal for Research in Mathematics Education*, *39*(4), 372–400.

Hill, H. C., Rowan, R., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *41*, 371–406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11–30.

Kilian, P., Glaesser, J., Loose, F., & Kelava, A. (2017). Structure of pedagogical content knowledge in maths teacher education – initial results of the Maths Teacher Education Study (MatTES). Manuscript submitted for publication.

Klein, F. (1908). *Elementarmathematik vom höheren Standpunkte aus. Teil 1: Arithmetik, Algebra, Analysis.* Teubner.

Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., ... Elsner, J. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 135 – 162). Waxmann.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, *100*(3), 716 – 725.

Krauss, S., Neubrand, M., Blum, W., & Baumert, J. (2008). The professional knowledge of german secondary mathematics teachers: Investigations in the context of the COACTIV project. In *11th International Congress on Mathematics Education (ICME 11).* Monterrey, Mexico.

Kunter, M., & Baumert, J. (2011). Das COCATIV-Forschungsprogramm zur Untersuchung professioneller Kompetenz von Lehrkräften - Zusammenfassung und Diskussion. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 345 – 366). Waxmann.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV.* Waxmann.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 151–173.

Löwen, K., Baumert, J., Kunter, M., Krauss, S., & Brunner, M. (2011). Methodische Grundlagen des Forschungsprogramms. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand

(Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 69 – 84). Waxmann.

Marks, R. (1990). Pedagogical content knowledge: From a mathematical case to a modified conception. *Journal of Teacher Education*, *41*(3), 3–11.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320-341.

McElvany, N., Schroeder, S., Baumert, J., Schnotz, W., Horz, H., & Ullrich, M. (2012). Cognitively demanding learning materials with texts and instructional pictures: teachers' diagnostic skills, pedagogical beliefs and motivation. *European Journal of Psychology of Education*, *27*(3), 403–420.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.

Mullis, I. V., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A., & Erberber, E. (2007). *TIMSS 2007 assessment frameworks.* Lynch School of Education: Boston College: TIMSS & PIRLS International Study Center.

Nathan, M. J., & Petrosino, A. (2003). Expert blind spot among preservice teachers. *American Educational Research Journal*, *40*(4), 905–928.

Ohle, A., & McElvany, N. (2015). Teachers' diagnostic competences and their practical relevance. special issue editorial. *Journal for Educational Research Online*, *7*(2), 5–10.

Petrou, M., & Goulding, M. (2011). Conceptualising teachers' mathematical knowledge in teaching. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 9–25). Dordrecht: Springer.

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667-696.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis: A Festschrift for Heinz Neudecker* (pp. 233–247). Boston, MA: Springer US.

Schoenfeld, A. H., & Kilpatrick, J. (2008). Toward a theory of proficiency in teaching mathematics. In D. Tirosh & T. Wood (Eds.), *International handbook of mathematics teacher education* (Vol. 2, pp. 321–354). Rotherdam, Netherlands: Sense Publishers.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review*, *57*(1), 1–23.

Tatto, M. T., Ingvarson, L., Schwille, J., Peck, R., Senk, S. L., & Rowley, G. (2008). *Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. conceptual framework.* East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. Rychen & L. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–66). Seattle, WA: Hogrefe & Huber Publishers.

Winsløw, C., & Grønbæk, N. (2014). The ecological Klein's double discontinuity revisited: contemporary challenges for universities preparing teachers to teach calculus. *RDM*, *34/1*, 59–86.

Chapter 2

Content knowledge and pedagogical content knowledge of trainee teachers for the academic track

The second topic of this thesis concerns prerequisites of students for the acquisition of mathematics content knowledge (MCK) and mathematics pedagogical content knowledge (MPCK) as well as further investigations about the relationship of those facets of knowledge. Thus, this chapter investigates determinants for knowledge. I show students' existing MCK and MPCK when they first enter university and the relationship between the two types of knowledge. Drawing on background information collected in addition to the tests, I show to what extend the knowledge states in MCK and MPCK can be explained by students' characteristics. Skills in the field of MCK show stronger dependence on covariates like performance measures from school than those of MPCK. The choice of the study program itself – teacher candidate (Bachelor of Education) or Bachelor of Science – and an affinity to teach, for which the choice of study program may be an indicator, does not lead to better performances on the MPCK test. See figure 2.1 for the integration of this chapter in the MatTES framework.

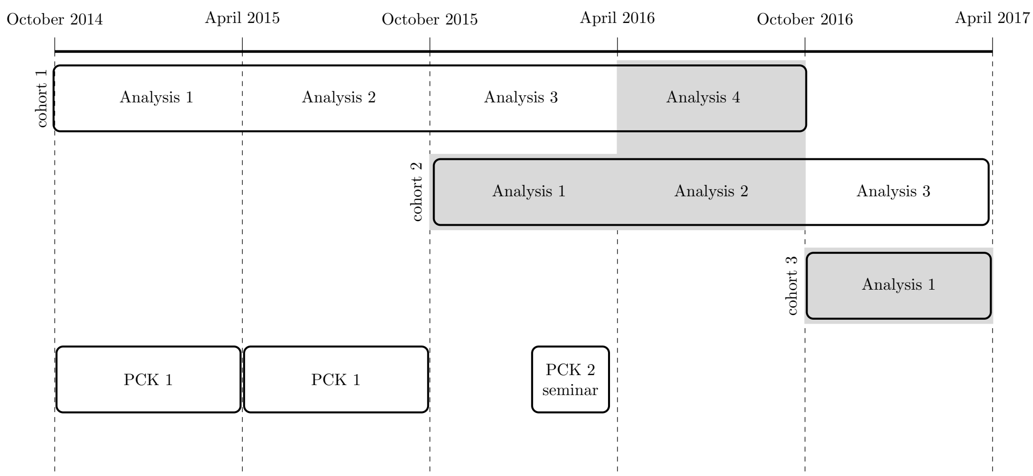

Figure 2.1: Integration of the study *content knowledge and pedagogical content knowledge of trainee teachers for the academic track* in the design of the Maths Teacher Education Study (MatTES). Highlighted lectures and seminars are part of the analysis.

Like the previous chapter, this chapter can be read in isolation and represents a stand-alone study. The theoretical background as well as sample and method information is provided.

Because of the relevance for German teacher education, the contents of this chapter have been submitted in German to a scientific journal (Glaesser, Kilian et al., submitted manuscript, 2018). Authors of the submitted article are: Judith Glaesser, Pascal Kilian, Christoff Hische, Jonathan Walz, Frank Loose and Augustin Kelava (in that oder). Christoff Hische and Jonathan Walz assisted in data collection. Frank Loose and Augustin Kelava had advisory functions and Judith Glaesser is the lead author of the submitted German article. The chapter was written on my responsibility.

2.1 Introduction

At present, skills and knowledge – in Germany frequently referred to as competencies – form a central topic in the area of eduction. Amongst other things, this raises questions about which skills teacher trainees already have at the outset of their studies and which ones should and could be taught at university. The focus of this chapter is again on content knowledge (CK) as well as on pedagogical content knowledge (PCK) with the goal of investigating the structure of knowledge on a conceptual and empirical level. For this purpose the knowledge of teacher trainees at the University of Tübingen was tested. Additionally covariates of the students were collected. The results of MatTES revealed a relation of CK and PCK. However the study also showed that it is possible to distinguish the two types of knowledge (see chapter 1). I showed that within PCK a further differentiation into different facets can be made.

The following chapter is structured as follows. First other studies and their theoretical competence models are presented. Second, the instruments and the sample is described. Followed by, third, the results and fourth, the discussion including implications for teachers education.

2.1.1 Theoretical background

Shulman (1986, 1987) suggested a possible differentiation of teachers' professional competence. According to him, professional knowledge includes CK and PCK beside the general pedagogical knowledge. Classroom management and organization are part of general pedagogical knowledge (Baumert & Kunter, 2011; König et al., 2017; Voss, Kunina-Habenicht, Hoehne, & Kunter, 2015). Despite the importance of those skills they will not be part of this study. The focus of this chapter is on CK and PCK. Those play a major role in literature, even though studies differ with regard to which other, additional competences and knowledge facets within teachers' professional competence they include in their investigations.

CK includes knowledge about contents and deep understanding of the subject matter. The acquisition of this knowledge is independent of the study program. In particular, it is not a type of knowledge confined to teachers. In contrast to CK, PCK is the specific knowledge needed by teachers to communicate subject matter to their students in school. In math this includes knowledge about the representation and explanation of mathematical contents as well as the knowledge about students' misunderstandings and how to avoid or adjust those.

Three big studies refer explicitly to Shulman and apply some version of his model. I refer to the studies of the Michigan-group including Deborah L. Ball (e.g. Ball, Lubienski, & Mewborn, 2001; Hill, Ball, & Schilling, 2008; Hill, Schilling, & Ball, 2004), the COACTIV study (Krauss et al., 2011) and TEDS-M (Teacher Education and Development Study in Mathematics – Blömeke, Kaiser, & Lehmann, 2010; Döhrmann, Kaiser, & Blömeke, 2012; Döhrmann, Kaiser, & Blömeke, 2010). Those studies introduce theoretical models which are then empirically tested. The authors of those study agree on several points. Throughout the studies the classification in CK and PCK is seen as reasonable and PCK as distinguishable from CK and from the general pedagogical knowledge albeit with overlap especially between CK and PCK. There is also a broad consensus that CK can be seen as a requirement for PCK to a certain degree.

Differences in the studies can be found concerning the operationalization of knowledge, further differentiations and the inclusion of further aspects of teachers' professional knowledge (e.g. motivation and beliefs). However, the authors note the conceptual similarity of their studies (Blömeke et al., 2010; Krauss et al., 2011).

2.1.2 Relationship of CK and PCK

The Michigan studies (e.g. Ball et al., 2001; Hill et al., 2008, 2004) showed a substantial overlap of CK and PCK. In constructing the items, it was assumed that two distinct dimensions exist (each including sub-dimensions) and the applied items cover all the postulated facets. Empirically the assumption of distinguishable dimensions was basically justified even though in the factor analysis some PCK items loaded on a factor which comprises CK items. Another factor analysis which allowed for simultaneous loading on multiple factors resulted in a large number of items loading on both the CK and the PCK factor. The authors concluded that both CK and PCK could be drawn on to answer those items (see Hill et al., 2008, p. 385).

The overlap between CK and PCK was also found in the TEDS studies (e.g. Döhrmann et al., 2012). The authors point out the challenges for measurement and empirical separation of CK and PCK implied by this overlap: “It is impossible to construct disjoint subdomains, because the solution of an item in the domain MPCK [mathematics pedagogical content knowledge] generally requires MCK [mathematics content knowledge]” (Döhrmann et al., 2012, p. 336). In the following I also use the more mathematical specific terms of *mathematics content knowledge* (MCK) and *mathematics pedagogical content knowledge* (MPCK) instead of CK and PCK. Those empirical challenges are expected because the possible overlap is inherent in the theoretical conceptualization which assumes the common existence of both aspects of knowledge, and teachers resort to them when teaching. It is assumed that those dimensions are not independent. Profound MCK is rather seen as a (necessary but insufficient) requirement for MPCK, also referred to by Döhrmann et al. (2012) in their discussion of example items. The same conclusion was found by Baumert and Kunter (2006): Content knowledge seems to be a necessary but insufficient requirement for high-quality teaching and for students’ progress. Content knowledge is the foundation on which didactic flexibility may emerge (original citation: “Fachwissen scheint eine notwendige, aber nicht hinreichende Bedingung für qualitätsvollen Unterricht und Lernfortschritte der Schülerinnen und Schüler zu sein. Fachwissen ist die Grundlage, auf der fachdidaktische Beweglichkeit entstehen kann.” [translated] (Baumert & Kunter, 2006, p. 496)). In order to act successfully from a didactic point of view, the teacher ideally has a big repertory of MCK which surpasses clearly, both in broadness and depth, the competences that should be acquired by the students (Krauss, Neubrand, et al., 2008). Profound MCK should ensure that the form of reasoning and the establishment of connections, which are involved in setting conceptual knowledge on a secure base, can be carried out in such a way that it can be deployed in the subject specific process of creating knowledge, here in mathematics (original citation: “ Argumentationsweisen und das Herstellen von Zusammenhängen, mithin das Sichern von begrifflichem Wissen, derart erfolgen kann, dass es an die typischen Wissensbildungsprozesse des Faches, hier der Mathematik, anschließen kann” [translated] (Krauss, Neubrand, et al., 2008, p 238)). The importance of MCK for MPCK appears in another result: Teacher education for the academic track includes more MCK but less MPCK in contrast to teacher education for other school types. Nevertheless teachers teaching on the academic track not only show better performances on MCK but also in MPCK compared to their colleagues of other school types. (Krauss, Neubrand, et al., 2008). A broad repertory of MCK facilitates the application of different approaches and representations, adjusted for students' needs, to enable a better and deeper understanding (Ball et al., 2001).

2.1.3 Structure and dimensions of MPCK

The Michigan studies (e.g. Ball et al., 2001; Hill et al., 2008, 2004) distinguish three facets of knowledge within MPCK: the knowledge of content and students (KCS), the knowledge of content and teaching (KCT) and the knowledge of curriculum. In COACTIV, the facets knowledge of mathematics-specific instructional strategies, knowledge of students' mathematical thinking and knowledge of mathematical tasks are examined as part of MPCK. Those facets were found in a confirmatory factor analysis (Krauss et al., 2011; Krauss, Brunner, et al., 2008). TEDS-M on the other hand assumes two dimensions for the structure of MPCK: curricular knowledge and planning for mathematics teaching and another interactive part entitled 'enacting mathematics for teaching and learning' (Döhrmann et al., 2012; Döhrmann et al., 2010; Tatto et al., 2008). For test construction within this study I developed a competence model which relies on the models summarized here. For the heuristic model I distinguished the two facets *instruction* and *diagnostic competence*.The model is summarized in figure 2.2.

Teaching relies on an interaction of both facets: Teachers prepare contents, present them and reflect in discussions in class how students process the contents and where problems may arise. Support in form of further instruction or alternative representations can be given if necessary as follow up. At the core of teaching there are teachers' reflection on their own actions and on potential modifications. This interaction can be represented as a circle, shown in figure 2.3.

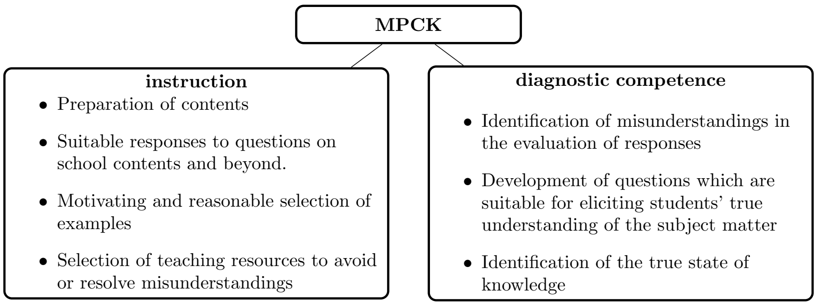

Figure 2.2: Two facets of MPCK

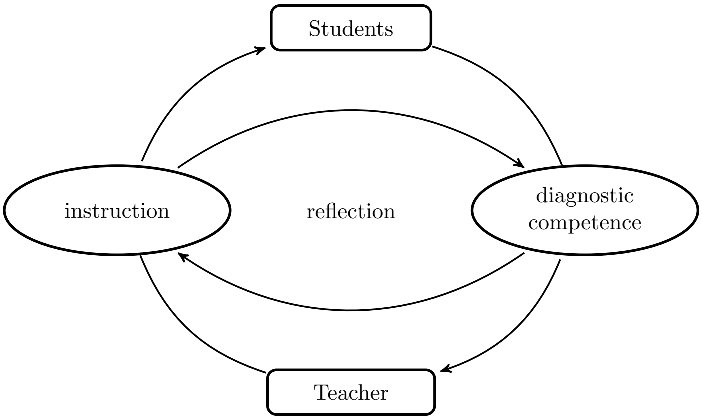

Figure 2.3: Cycle of teacher-student interaction

2.1.4 Research question

Theories and results so far demonstrate the importance of both MCK and MPCK for teaching. In addition, in this chapter I consider the question whether, and if so how strongly the facets are related for students in the initial phase at university. Because of the fact that a relationship between the two types of knowledge may be expected, I am interested in the prerequisites teacher candidates have at the outset of their university course for the acquisition of those facets. Potential important factors for the performance on the MCK and MPCK tests developed by the project team are the GPA (final grade point average from school), the math grade in the final exam in school, sex, the school type in which the final exam was taken (general-education

Gymnasium or other) as well as the study program (here the teacher candidates studying for a Bachelor of Education (B.Ed.) degree and the other Bachelor of Science (B.Sc.) students). Those variables were chosen on the basis of the following considerations. MCK is already taught in school, thus an influence of school performance measures is expected, especially in mathematics. This is not directly the case for MPCK but because of the fact that, as mentioned, a relationship with MCK may be expected, it is plausible to expect a connection between good performances in school and good performances in MPCK. In the framework of this study there is no expectation on the potential effect of students's sex and of school type, but the analysis of those variables is interesting nevertheless. Girls and women are a minority in the STEM fields (science, technology, engineering and mathematics)[1] so an empirical examination of differences in sex could be informative. Note that in this study female students are a minority in the B.Sc. study programs but this does not apply for the B.Ed. study program in mathematics. It could be informative, both for individuals as well as policy-makers, to investigate a potential influence of the school type. Both groups should be able to rely on sound empirical results concerning the characteristics of different school types. Because the special interest of this study is teacher candidates' acquisition of competences I investigate the influence of the study program even though in the initial phase at university, during which this study is conducted, the schedules of the study programs do not differ. In particular, this means that teacher candidates did not attend lectures for MPCK at this time. Therefor, a difference in performance in the MPCK domain is not to be expected based on degree program. Nonetheless, it is conceivable that the choice of the B.Ed. itself – and an affinity to teach for which the choice of study program may be an indicator – does lead to a better performance in the MPCK test. In the next section I introduce the sample, the variables and the tests employed.

[1]Comparable with the MINT disciplines (mathematics, informatics, science and technology) in Germany

2.2 Methods

2.2.1 Study design and sample

The database for this study consists of three cohorts who had commenced their studies in winter semester 2014/15 (in the following referred to as cohort 1), winter semester 2015/16 (cohort 2) and winter semester 2016/17 (cohort 3). Because the tests employed here are under continuous development as part of the project I was not able to use all cohorts for all investigations. The MCK test was carried out in cohort 2 and cohort 3 (in their respective first semester 2015/16 and 2016/17). The MPCK test was carried out in cohort 1 and cohort 2 in the summer semester 2016. Therefore, cohort 2 appears in all the analyses, but because the MPCK test was carried out in their second semester the sample is much smaller at that time, due to dropouts during and after the first semester. This leads to different samples sizes for the MCK and the MPCK test (see table 2.1). I refer to the subsample within cohort 2 which attended in both test as cohort 2*. All of the cohorts include both B.Sc. and B.Ed. students and their participation in the study was requested when they attended the analysis lectures. The participation was voluntary and anonymous.

2.2.2 Background variables (dependent variables)

Table 2.1 summarizes the distribution of the background factors in all the cohorts.

Columns one and two – cohort 1 and cohort 2* – describe students included in the MPCK analysis, columns three and four, those included in the MCK analysis. It can be seen that the cohorts used within each set of analyses do not differ substantially, but there are differences between the combined groups (cohort 1 + 2* and cohort 2 + 3). School performances of cohort 1 + 2* are better on average than those of cohort 2 + 3. The reason for this relates to the point in time at which the tests were carried out. Cohort 2 + 3 includes all students at the beginning of a mathematical study program at university whereas cohorts 1 + 2* were more highly selected, for example due to voluntary and involuntary dropouts. Due to the relationship between performance measures and dropouts (see chapter 3) students who had shown higher performance in school remain in the cohorts in the second semester.

Table 2.1: Background variables of the cohorts

		MPCK		MCK	
cohort		cohort 1	cohort 2*[a]	cohort 2	cohort 3
lecture		Analysis 4	Analysis 2	Analysis 1	Analysis 1
sex	male	46 (54.1%)	47 (48.0%)	123 (57.7%)	106 (58.6%)
	female	39 (45.9%	51 (52.0%)	90 (42.3%)	75 (41.4%)
school type	general ed.	74 (87.1%)	88 (89.8%)	164 (77.0%)	150 (82.9%)
	other	11 (12.9%)	10 (10.2%)	49 (23.0%)	31 (17.1%)
study program	teacher	42 (49.4%)	66 (67.3%)	105 (49.3%)	76 (42.0%)
	other	43 (50.5%)	32 (32.7%)	108 (50.7%)	105 (58.0%)
GPA	mean	1.8	1.8	2.0	2.1
	min	1.0	1.0	1.0	1.0
	max	3.5	3.4	3.6	3.4
math grade	mean	12.8	12.3	11.6	11.5
	min	4	5	3	1
	max	15	15	15	15
N		85	98	213	181

Note. [a] Subgroup of cohort 2 which participated in the MPCK test
school type = general-education Gymnasium or not; GPA = grade point average in the final exam in school. Scores range from 1 to 6 with 1 as the best score; math grade = math grade in the final exam in school. Scores range from 1 to 15 with 15 as the maximum score

2.2.3 MCK and MPCK tests (independent variables)

I developed test for MCK and MPCK. In the following part I describe those tests and their development.

MCK test

For the MCK test I used a combination of TIMSS items (Baumert et al., 1999)[2] and items I had developed myself. The latter were developed in close cooperation with experts and lecturers involved in the Analysis 1 lecture. The test consisted of 13 dichotomous items, i.e. the maximum total sum score was 13 points (M = 6.9, SD = 2.3). Since students attending the analysis lectures were the participants of this study, most of the items can be found in that field. Compared to the TIMSS items my own items require a higher level of abstraction compared to school contents and are closer to the contents of the Analysis 1 lecture. Since the TIMSS items had been developed for a broader sample of students in school, my participants of students in a mathematical study program performed rather well. The TIMSS items and the items I had developed myself showed a correlation of r = 0.28. In appendix A.2 an example of a task developed by me is given.

MPCK test

The test for MPCK consists of items from TEDS-sM (Buchholtz et al., 2012)[3] and of items developed by myself. Again, experts in both MCK and MPCK were involved in the development. This collaboration contributed to content validity. In test construction I attached great importance to obtaining a test suited to the students' presumed state of knowledge. This involved, amongst other things, to refrain from the use of (at this point unknown) technical terms in both MCK and MPCK. For validity examination, the correlation of r = 0.32 between the TEDS-sM items and the MatTES items was calculated. This correlation points to similar underlying concepts but also suggests that my test items cover additional facets. Therefore, an extension of TEDS-sM seems justified. The highest possible test score on the MPCK test is 32 (M = 25.0, SD = 3.7). An example of a MatTES task developed by me can be found in appendix A.1.

[2] I used the tasks K5, K6, K4, L6 and L5 because of their fit to the contents of the Analysis 1 lecture

[3] I used the tasks SUG2_2, SUG2_3, SUG2_4, DS29_1, DS29_2, DS29_3, DBJ4_1, DBJ4_2 and DBJ4_4

2.3 Results

2.3.1 Relationship of MCK and MPCK

To examine how closely the dimensions of MCK and MPCK are related I calculated the correlation of the respective test performances. Note that in the previous chapter 1 (and in Kilian, Glaesser, Loose & Kelava (submitted manuscript, 2017)) I already showed that the dimensions can be separated empirically. Because only the participants of cohort 2 participated in both tests, only those students were involved in this calculation. The correlation was $r = 0.25$ ($n = 94$, $p<0.01$). This shows the expected relationship of MCK and MPCK. On the other hand the correlation is not so high as to support a claim of there being no difference between the dimensions. Therefore, it seems to be appropriate to separate the dimensions analytically and empirically.

2.3.2 Conditional factors for MCK and MPCK

In this part I investigated which background variables underpin good performances on the tests. For this purpose I carried out two linear regression analyses with the respective test performance as dependent variable and the background variables as independent variables. In table 2.2 I present the results for the MCK test. A few students participated twice in the MCK test because they repeated the Analysis 1 lecture for various reasons. If this was the case, I drew on the results from the first time the person had participated.

Sex was coded with 1 = female and 2 = male thus the positive coefficient indicates a better performance of the male students. A positive effect can also be seen for students who did their final exam in a general-education Gymnasium. As expected the test performance increases with better GPAs and math grades. The study program showed no significant coefficient.

I used the same independent variables for the analysis of the MPCK test. Additionally I was able to examine a potential effect of experience, because in cohort 1 the test was carried out in the fourth semester, whereas in cohort 2, the test was carried out in their second semester. This potential effect would

Table 2.2: Results of the linear regression for the MCK test

	Estimate	Std. Error	std. Estimate	t value	Pr(> $\|t\|$)
(Intercept)	4.186	0.947		4.421	0.000
sex	**0.538**	**0.224**	**0.113**	**2.396**	**0.017**
school type	**0.953**	**0.250**	**0.164**	**3.807**	**0.000**
GPA	**-0.894**	**0.228**	**-0.232**	**-3.924**	**0.000**
math grade	**0.266**	**0.048**	**0.319**	**5.592**	**0.000**
study program	-0.271	0.211	-0.058	-1.279	0.202

Note. **bold**: p < .5
N = 394 (cohort 2: n = 213, cohort 3: n= 181); adj. R^2: 0.290
school type = general-education Gymnasium or not; GPA = grade point average in the final exam in school. Scores range from 1 to 6 with 1 as the best score; math grade = math grade in the final exam in school. Scores range from 1 to 15 with 15 as the maximum score; study program = B.Ed. (teacher candidate) or B.Sc.

be the result of having experienced more teaching of math content knowledge because no lectures or seminars in the field of MPCK (for teacher candidates) were provided in this time period. Math content experience was coded in the regression model via the cohort membership (cohort 1 = 1), thus a positive effect means better test performances for more experienced students. In table 2.3 results of the regression analysis for the MPCK test are presented.

The sample size of cohort 2* differs from that reported for the MCK test (cohort 2). As mentioned before this is due to dropouts during and after the first semester. The MPCK test was carried out with the remaining students in cohort 2.

The analysis of the MPCK test showed similar relations as in the MCK test with the exception of the school type. The type of the school, where the final exam had been taken, did not have any relationship with the MPCK test performance. Sex was only just statistically significant but again with male students showing the better performance.

The cohort membership – which is a proxy for experience – had an effect which was only just statistically significant. Students in their fourth semester (second year of study) performed better than their colleagues in the second semester. This is remarkable in the framework of the MPCK investigation because the only additional experience gained during this part of the course

Table 2.3: Results of the linear regression for the MPCK test

	Estimate	Std. Error	std. Estimate	t value	Pr(> $\|t\|$)
(Intercept)	20.230	2.709		7.468	0.000
sex	**1.125**	**0.533**	**0.153**	**2.110**	**0.036**
school type	1.050	0.795	0.091	1.321	0.188
GPA	**-1.501**	**0.631**	**-0.228**	**-2.379**	**0.018**
math grade	**0.337**	**0.148**	**0.219**	**2.271**	**0.024**
study program	0.192	0.535	0.026	0.358	0.720
cohort 1	**1.149**	**0.500**	**0.156**	**2.298**	**0.023**

Note. **bold**: p < .5
N = 183 (cohort 1: n = 85, cohort 2*: n= 98); adj. R^2: 0.202
school type = general-education Gymnasium or not; GPA = grade point average in the final exam in school. Scores range from 1 to 6 with 1 as the best score; math grade = math grade in the final exam in school. Scores range from 1 to 15 with 15 as the maximum score; study program = B.Ed. (teacher candidate) or B.Sc.; cohort 1 = test in the fourth semester, compared to cohort 2 with the test in the second semester

of study were the mathematical lectures. No lectures or seminars in the field of MPCK were provided and thus attended. This result points to the relevance of MCK as a requirement for MPCK.[4]

Here too, the study program showed no significant effect on the test performance. On the one hand, this is not surprising because as mentioned before at the time of the study the teacher candidates had not attended any MPCK lectures, on the other hand, this shows that the decision to become a teacher and any affinity to teaching indicated by this decision does not lead to a better performance in MPCK.

The samples for the analysis in MCK and MPCK differed in composition (i.e. cohort membership) as well as in sample size. For this reason, any conclusion drawn on the basis of their comparison should be viewed with caution. Nevertheless I will undertake such a comparison to examine the two facets of competence and knowledge. The coefficient R^2 was higher in the analysis of MCK than in the analysis of MPCK. This suggests that the variables employed in the analyses are more strongly related to MCK test

[4]The analysis was repeated without the cohort membership to use the same set of variables as in the MCK linear regression. For that model I achieved $R^2 = 0.183$. There was no change in direction and significance of the other effects.

performance than MPCK test performance. Two of those variables – GPA and math grade – can be assumed to be indicators of conventional content knowledge and cognitive abilities. Those in turn are more closely related to the competences and skills measured by the MCK test than to those measured by the MPCK test. The latter represents competences and skills which are demanded by the teacher in everyday school life. Thus performance on this test depends less strong on school math skills and general cognitive abilities than performance on the MCK test.

2.4 Discussion

As noted in the introduction, both based on theoretical considerations and previous empirical studies, it seems reasonable to assume that MCK and MPCK occur together and that sound MCK is a prerequisite for MPCK. This could be confirmed: the correlation between the two test performances points to a relationship. At the same time, the correlation of r = 0.25 is low enough to indicate different and distinguishable dimensions.

Accordingly, conditional factors for the performance in those dimensions were evaluated separately for both test performances. Again there are parallels: for both dimensions the variables sex (with better performance on the part of male students) and performance measures from school – the GPA and the math grade (with better grades in school leading to better performance) – showed statistically significant effects. Furthermore, for MCK, having taken the final exam at a general education Gymnasium was a factor associated with higher test performance. This was not the case for the MPCK performance. The study program showed no effect on test performance which means that no differences with regard to test performance were found between teacher candidates (B.Ed.) and their colleagues (B.Sc.). Using the cohort membership, for MPCK it was possible to measure the effect of experience in terms of the duration of the mathematical education (this means having been taught MCK but not MPCK). This analysis showed a positive effect of mathematical experience on both test performances. I discuss those results in relation to two aspects. The results are important, firstly, for the question concerning the relationship between MCK and MPCK and, secondly, they concern various individual conditions for the acquisition of MCK and MPCK.

2.4.1 Relationship of MCK and MPCK

Both the correlation of the dimensions as well as the similarity of the importance of the factors point to the existence of the relationship which had been expected on the basis of theory. Even though knowledge from school – operationalized by GPA and math grade – does not include MPCK components, better performance in school is not only associated with better performance on the MCK test but also with better performance on the MPCK test. Mathematical skills appear to be helpful in completing tasks relating to MPCK. In addition to the math grade, students' GPA showed a positive effect. This can be interpreted as pointing to the importance of general cognitive abilities.

Another indicator for the importance of subject matter education offers the effect of cohort membership on MPCK test performance. No MPCK education had been received at the time of the study but the participants in cohort 1 were more advanced by one year in their mathematical education compared to the participants of cohort 2 and performed better on the MPCK test. This leads to the assumption that the subject matter contents were helpful for solving questions in the MPCK field. These results have implications for teacher education: the importance of sound MCK for MPCK was confirmed. This stresses once again the importance of subject matter education for teacher candidates and contradicts the idea of a simplified study program for the Bachelor of Education students compared to their Bachelor of Science colleagues.

2.4.2 Preconditions for the acquisition of MCK and MPCK

The importance of cognitive abilities for good MCK and MPCK performance has been mentioned above. Thus in this part I focus on the remaining variables sex, school type and the study program.

The data does not allow for explanations of the effect of students' sex. Similar results were found in Blömeke (2013): performance by novice Englisch and German teachers were independent of the participants' sex, whereas on mathematical tests there was a relationship with male participants performing better, as in the present study. Reasons underlying this relationship are thought to be socio-cultural conditions, in other words less support, lower expectations and fewer formal and informal opportunities to learn mathematical contents for females (Blömeke, 2013).

Likewise, it is not possible to draw any firm conclusions concerning the reasons underlying the effects of school type which were shown in the present study. It is conceivable that participants who did not take their final exam in a general education Gymnasium did not attend the academic track (Gymnasium) at all until upper secondary school. Less demanding mathematical requirements and depth compared to the academic track may have led to deficiencies which may not have become apparent in the final math grades or the GPA but in the general mathematical skills. This assumption should be examined with appropriate data.

The non-existent effect of the study program (B.Ed. compared to B.Sc.) is interesting because it might have been assumed that the choice of the study program B.Ed. is linked to an affinity for teaching and thus for pedagogical contents. Nevertheless the teacher candidates showed no better performances on the MPCK test than their B.Sc. colleagues which also means that the B.Sc. students have the same conditions required for becoming a teacher, at least according to the background information included in this study. Given the shortage of teachers in STEM subjects, this result could be important since it means that B.SC. students may be open to suggestions to switch to the B.Ed. study program once they realize that they have the necessary characteristics.

2.4.3 Outlook

The results reported here were obtained during the early stages of the research project. A range of questions follows which are to be addressed in subsequent stages. Firstly, the tests are continuously revised and the samples expanded. This involves further examination of the structure especially of MPCK (first results are reported in chapter 1 or in Kilian et al. (submitted manuscript, 2017)). Secondly, the question of whether the effects reported here are still in evidence during later stages of the program of study will have to be examined. If this were not to be the case, this could indicate that the university can contribute to narrowing the gap between students. If it did turn out to be the case, various conclusions can be drawn. First, this could serve to underscore the importance of improving of teaching at university to support weaker students. Second, school performance measures could be considered as criteria for admission restrictions for mathematical study programs (for more information on this topic see chapter 3). Finally an expansion to the second phase of teacher education is planned to examine the competences and skills examined here and how they are applied in practice.

References

Ball, D. L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (Vol. 4, pp. 433 – 456). New York, NY: Macmillan.

Baumert, J., et al. (Eds.). (1999). *Testaufgaben zu TIMSS/III: Mathematisch-naturwissenschaftliche Grundbildung und voruniversitäre Mathematik und Physik der Abschlussklassen der Sekundarstufe II (Population 3)* (Vol. 62). Berlin: Max-Planck-Institut für Bildungsforschung.

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, *9*(4), 469–520.

Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 29 – 53). Waxmann.

Blömeke, S. (2013). Einleitung: Professionelle kompetenzen im studienverlauf [Book Section]. In S. Blömeke et al. (Eds.), *Professionelle kompetenzen im studienverlauf. weitere ergebnisse zur deutsch-, englisch und mathematiklehrerausbildung aus teds-lt* (p. 7-24). Münster: Waxmann.

Blömeke, S., Kaiser, G., & Lehmann, R. H. (2010). Teds-m 2008 sekundarstufe i: Ziele, untersuchungsanlage und zentrale ergebnisse [Book Section]. In S. Blömeke, G. Kaiser, & R. H. Lehmann (Eds.), *Professionelle kompetenz und lerngelegenheiten angehender mathematiklehrkräfte für die sekundarstufe i im internationalen vergleich* (p. 11-

37). Münster: Waxmann.

Buchholtz, N., Scheiner, T., Döhrmann, M., Suhl, U., Kaiser, G., & Blömeke, S. (2012). *TEDS-shortM: Kurzfassung der mathematischen und mathematikdidaktischen Testinstrumente aus TEDS-M, TEDS-LT und TEDS-Telekom.* Hamburg: Universität.

Döhrmann, M., Kaiser, G., & Blömeke, S. (2012). The conceptualisation of mathematics competencies in the international teacher education study TEDS-M. *ZDM*, *44*(3), 325–340.

Döhrmann, M., Kaiser, G., & Blömeke, S. (2010). Messung des mathematischen und mathematikdidaktischen wissens: Theoretischer rahmen und teststruktur [Book Section]. In S. Blömeke, G. Kaiser, & R. H. Lehmann (Eds.), *Professionelle kompetenz und lerngelegenheiten angehender mathematiklehrkräfte für die sekundarstufe i im internationalen vergleich* (p. 169-196). Münster: Waxmann.

Glaesser, J., Kilian, P., Hische, C., Walz, J., Loose, F., & Kelava, A. (2018). Fachwissenschaftliche und fachdidaktische Kompetenzen bei angehenden Gymnasiallehrer/innen der Mathematik. [Content knowledge and pedagogical content knowledge of trainee teachers for the academic track]. Manuscript submitted for publication.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: conceptualizing and measuring teachers' topic-specific knowledge of students [Journal Article]. *Journal for Research in Mathematics Education*, *39*(4), 372-400.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11–30. Retrieved from `http://dx.doi.org/10.1086/428763` doi: 10.1086/428763

Kilian, P., Glaesser, J., Loose, F., & Kelava, A. (2017). Structure of pedagogical content knowledge in maths teacher education – initial results of the Maths Teacher Education Study (MatTES). Manuscript submitted for publication.

Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., ... Elsner, J. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 135 – 162). Waxmann.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, *100*(3), 716 – 725.

Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M., & Jordan, A. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und-Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik*, *29*(3-4), 233–258.

König, J., Doll, J., Buchholtz, N., Förster, S., Kaspar, K., Rühl, A.-M., … Kaiser, G. (2017). Pädagogisches Wissen versus fachdidaktisches Wissen? Struktur des professionellen Wissens bei angehenden Deutsch-, Englisch- und Mathematiklehrkräften im Studium. *Zeitschrift für Erziehungswissenschaft*.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review*, *57*(1), 1–23.

Tatto, M. T., Ingvarson, L., Schwille, J., Peck, R., Senk, S. L., & Rowley, G. (2008). *Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. conceptual framework.* East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.

Voss, T., Kunina-Habenicht, O., Hoehne, V., & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft*, *18*, 187-223.

Chapter 3

Identification of risk factors and prediction accuracies for math students' success in a first semester math lecture

In this chapter I focus solely on the acquisition of MCK in the sense of a successful start of studies. The importance of MCK for teaching (e.g. Ball, Lubienski, & Mewborn, 2001; Krauss, Neubrand, Blum, & Baumert, 2008) and as a prerequisite for MPCK (Baumert & Kunter, 2011) has been shown in former studies. Instead of examining students' mathematical performance and knowledge directly, in this chapter I investigate the success and possible dropout of students in a first semester mathematical lecture. I chose the mandatory Analysis 1 lecture as reference point for the start of studies. This study connects a number of prerequisites of students at the start of studies and their relation to success or dropout. The frame of reference within the MatTES project is shown in figure 3.1.

As teachers' MCK is so important, dropout rates in these courses can provide valuable information. Those lectures are responsible for the acquisition of this knowledge and thus can serve for investigations about the role of different prerequisites. Investigations about risk groups concerning success

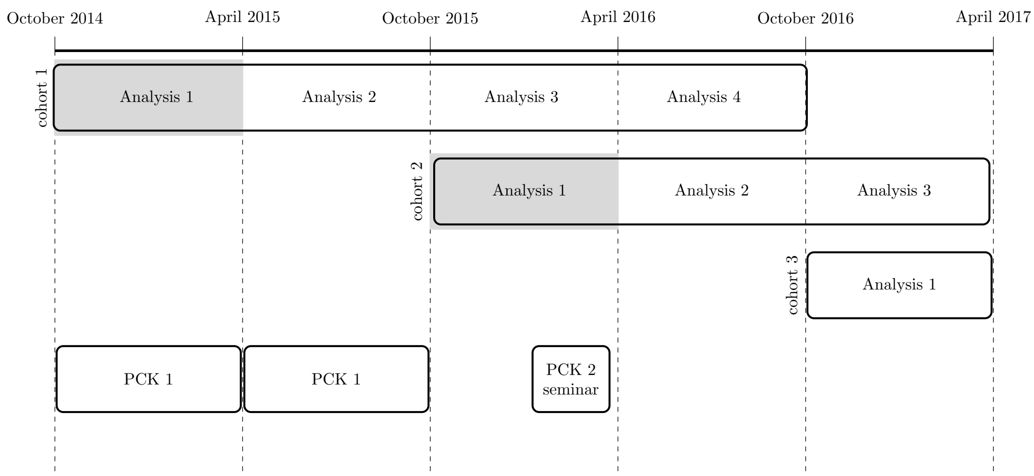

Figure 3.1: Integration of the study *identification of risk factors and prediction accuracies for math student's success in a first semester math lecture* in the design of the Maths Teacher Education Study (MatTES). Highlighted lectures and seminars are part of the analysis.

can reveal important information about the group of teacher candidates. As well as previous chapters, this chapter can be read as a separate chapter (without the context of the whole paper).

3.1 Introduction

High dropout rates in mathematics and general in the STEM fields (science, technology, engineering and mathematics) – more precisely in Germany the so called MINT disciplines (mathematics, informatics, science and technology) – are not a new phenomenon. According to Heublein (2014), in Germany 39% of the students in MINT disciplines drop out of the Bachelor program. Compared to the german general average of 33% this is a rather high dropout rate. In mathematics the dropout rate is even higher with 47%. These numbers solely include real drop outs, university transitions or changes in study program are not included. Those rates are not to be confused with shrinkage rates, which refer only to the dropout within one discipline at one university and can be defined for study programs but not on an individual level. Therefore the shrinkage rate at universities is even higher. In contrast to the MINT dropout rates during the bachelor, the 5% dropout rate during the master is much lower (Heublein, 2014). This focuses the investigation on the initial phase of the studies.

The investigation of these unusually high dropout rates in MINT disciplines becomes even more interesting and relevant knowing that those students show high cognitive prerequisites (Nagy, 2007). Nagy showed high correlations of cognitive abilities with realistic and investigative vocational interests (based on the vocational interest model of Holland (1997)) where those disciplines are classified.

3.1.1 Structure of this chapter

First, in the following section I briefly discuss different frameworks and results of former studies and summarize the current state of research. Secondly, I introduce studies more specific to the topic of this chapter, followed by thirdly, the design of the present study and the research questions. The methods and results sections follow, and I close with a discussion of our findings.

3.1.2 Current state of dropout research

The investigation of dropouts can draw on a high quantity of studies and literature. In terms of dropout related factors a big variety of variables have been investigated. Variables have been identified on individual, institutional, environmental-related and system-related levels. Following Bean (2005) and Burrus et al. (2013) variables can be classified in (a) institutional environment factors, (b) student demographic characteristics, (c) commitment, (d) academic preparation and success factors, (e) psychosocial and study skills factors, (f) integration and fit, (g) student finances and (h) external pull factors.

In the following I summarize a few results of former studies. I focus on factors within (b) student demographic characteristics and (d) academic preparation and success factors, as these reflect the design of this study best. Examples for student demographic characteristics are, among others, the student's age, sex, race and ethnicity. Feldman (1993) reveals a non linear connection between age and student dropout – younger and older students showed higher dropout probabilities. Hagedorn, Maxwell, and Hampton (2001) show a slightly negative effect of age on students' retention. For student's sex no simple linear effects have been described, but interactions with other variables have been shown (e.g. with the existence of children in Leppel (2002)). Academic abilities like school performance measures and general performance measures are examples of the (d) academic preparation and success factors. Standardized tests show the expected high correlation with student's retention at university (e.g. Bean, 2005).

In summary, it can be said that a wide range of variables show a correlation with student dropout, indicating a complex relationship wherein the effect might rely on combinations of variables. This gives motivation to expand the approaches from linear models to models which take this complexity into account, for example in prediction models.

The dropout literature relies mostly on two frameworks for understanding students' dropout decisions. First there is Tinto's *theory of student departure* (Tinto, 1975, 1987). A central point of his theory is the students' integration and interaction with the faculty, staff and peers in both academic and social settings (Burrus et al., 2013; Tinto, 1993). The model has been validated and generally adjudged to be a useful framework (e.g. Terenzini & Pascarella, 1980). The second framework is the *model of student attrition* by Bean (1980, 1983, 2005). This framework implies more external factors, e.g. the expenditure of time, financial resources or the students' responsibility for their families. In both models the intention to drop out is explained by the variables: contentment with the study program, the pursuit of the final degree and the power of endurance.

More recent approaches focus more on individual characteristics of the students. Schiefele, Streblow, and Brinkmann (2007) for example, show that differences between students who dropout and persistent students are mostly found in motivation, social competence, perceived teaching quality, the self-evaluated knowledge state and the use of learning strategies.

3.1.3 Prerequisites of math students and prediction of the success in the first semester mathematical lecture Analysis 1

Several studies, especially for mathematics, have been conducted with focus on the transition from school to university including individual prerequisites. Those studies include investigations about the social and institutional context of math (Gueudet, 2008), reasons for problems at the transition to university (Heublein, Hutzsch, Schreiber, Sommer, & Besuch, 2009), investigations about the differences between math in school and at universities from the lecturers point of view (Grünwald, Kossow, Sauerbier, & Klymchuk, 2004). Specific for teacher candidates, the positive correlation between the growth in competence of students and their previous knowledge, the GPA and the interest in the teacher study program (Eilerts, 2009). Parts of the risk factors of those studies can be summarized as student's prerequisites prior to university. Below I discuss some of those prerequisites with special regards to teacher candidates.

Prerequisites

Regarding the choice of the study program, it seems to be a public opinion that there is a negative selection within the mathematics students towards the teacher candidates. It is assumed that weaker students choose the teacher program Bachelor of Education (hereafter referred to as B.Ed.) instead of the pure math program Bachelor of Science (referred to as B.Sc.). External reasons like occupational safety and longer vacation periods are named motivations for this choice. These reasons might be more important than the motivation of becoming a teacher itself (Blömeke, 2005; Klusmann, Trautwein, Lüdtke, Kunter, & Baumert, 2009).

However, a negative selection was not found by Klusmann et al. (2009), comparing school grades, cognitive abilities and results of a standardized math test (Third International Mathematics and Science Study (TIMSS, e.g. Baumert et al., 1999)), between teacher candidates for the academic track (B.Ed.) and non teacher candidates at university (B.Sc.). Those results refer only to measurements at the end of school and give no information about the success at the university. In their study Klusmann et al. (2009) used data of the study *Transformation des Sekundarschulsystems und akademische Karrieren* (TOSCA; Köller, Watermann, Trautwein, & Lüdtke, 2004), which does not differ between B.Ed. and B.Sc. students. Besides sex, the analysis was controlled for the field of the study subject (inclusion of at least one subject within the field of science or not). Therefore the dataset is not differential enough for statements within the subject of mathematics. Even though the initial prerequisites might be the same, both study programs start to differ vastly already in the first semester. Teacher candidates have to face the double pressure of two majors, differences in success and dropout rates during the semester might occur. Due to the importance of content knowledge for teaching and its relation to pedagogical content knowledge (e.g. Kunter et al., 2011) it is a critical question if the teacher candidates fall behind their colleagues in terms of success and dropout rates already in the first semester.

3.1.4 The Present Research

In this study I analyze the success in the first semester lecture Analysis 1. I compare the prerequisites of students in different study programs, attending this lecture and connect it to the above mentioned studies (e.g. Klusmann et al., 2009). These prerequisites are then used to predict dropouts in this time period. For a more detailed definition of the term dropout in this study I refer to the methods section below.

As seen in former studies and frameworks (e.g. Bean, 1980, 1983, 2005; Burrus et al., 2013; Schiefele et al., 2007; Tinto, 1975, 1987, 1993), a broad variety of possible dropout predictors or risk factors can be tested. Even though the different models show overlaps in the sets of risk factors, some are contradictory. This shows the multilateral structure of student dropout and the complex relation of different risk factors. In this study I use a very small set of possible predictors by only including, (a) data collected at the beginning of the semester (excluding for example performance measures during the semester or different state variables), and (b) variables which can be collected with little cost at the beginning of the lecture. This variable set includes for example personal data like age and sex, as well as performance measures from school and an initial math test. The reasons for this choice are (a) to find risk factors in students' prerequisites excluding the student's behavior during the lecture (e.g. the expenditure of time) and (b) to enable universities to use the results of this study with little cost. Possible actions could be entrance qualifications and the identification of risk groups early in the program to provide support courses and interventions. The standard method to identify significant risk variables, for a binary variable like student dropout, is the logistic regression. As the logistic regression only works on linear relations, we will use prediction models that are able to take complex interactions into account. This approach aims to be more practical compared to the theoretical approaches discussed above. As shown, risk factors for student dropout can be found in students' personality, school backgrounds, social and academic integration and many more. Although those broad theoretical approaches are important for the understanding of dropout risk factors, they are not of practical use for universities (in most cases simply because of the inaccessibility of the wide range of variables). In order to enable universities to quickly identify risk groups the focus of this study is the prediction of student dropout using only a few, leviable variables, in sophisticated models.

Research questions

The research questions of this study can be summarized in four parts. First I investigate the question regarding differences in the prerequisites of the students at the beginning of university, with special interest in teacher candidates (B.Ed.). The second questions concerns the identification and selection of variables – within the initially chosen set – which are most important for accurate dropout predictions. Important variables are either those with significant coefficients or those, selected by method specific variable selection algorithms. Third, in the prediction context, I identify an upper bound of prediction accuracy that can be achieved using only students' characteristics at the beginning of the lecture (again, no behavior measures during the lecture are included). The fourth research question concerns the interpretation and application of the prediction results. I investigate if the prediction results allow the identification of rules which assign students to risk groups depending on their prerequisites in terms of the used features.

3.2 Methods

3.2.1 Study design and sample

In this study I examine the Analysis 1 lecture of cohort 1 and cohort 2 of the MatTES dataset. In cohort 1 the study was conducted in the winter semester 2014/15 at the University of Tübingen. The Analysis 1 lecture for the second cohort took place in the winter semester 2015/16 (see figure 3.1).

Due to the curriculum, it is mandatory to participate in the Analysis 1 course in the first semester, for both B.Sc. and state examination (teacher training, B.Ed.) students. In Tübingen, the physics (B.Sc.) students participate in this course in the first semester as well. The schedule of the Analysis 1 lecture is comparable for both cohorts. Additionally to attending the lectures, students are divided into small tutorial groups.

Within the tutorial groups, students have to submit homework every week. The homework addresses problem sets introduced in the lecture and is discussed and graded in the tutorials after submission. In order to obtain the admission to the final exam, achievements in the context of those tutorials and on the problem sets is relevant. For cohort 1 and cohort 2 achievement goals differed. In cohort 1, the required achievement was to score 50% of the total available points on all the problem sets. For the achievement goal in cohort 2 a combination of a minimum total point score on the problem sets and the results of two tests during the semester were necessary. Those requirements were defined by the lecturers of the respective lectures.

In the analysis I combine both cohorts to reduce the dependence on lecturers and lecture schedules and gain more general results. I refer to the combined data set as 'dataC'.

Exclusions

Several participant in the cohorts are excluded from the analysis. First, students which already gained the admission for the final exam in Analysis 1 in a previous semester but did not take part or pass the exam are allowed to participate in the exams without achieving the admission again. Those students are excluded because the research questions refer to students, which participate in the whole program of the lecture. Note that as a results of this exclusion we can say that there are no students in the cohorts which actively participated in a former Analysis 1 lecture, even though it might not be their first mathematical semester. Second, in the analysis I only consider students, which major in mathematics (B.Sc), physics (B.Sc) or are teacher training students (B.Ed). This excludes a few students which participate voluntarly. Third, students are excluded if information on either their questionnaire, or the lecture is missing. About 95% of the students filled in the initial questionnaire. In order to be able to access students' lecture data an additional permission was needed. 95% of the students granted this permission.

3.2.2 Instrument

The dataset contains the results of every student on every problem set, which allows to follow the students development during the first semester. Additionally I can see the exact week of the dropout, if students quit. The open response problems were graded by the instructors of the tutorials. A high value is set on grading the items equally for all examinees. Additionally the results of the final exams are analyzed. Those results indicate success or failure of the Analysis 1 lecture. For further information about the preconditions of students, like personal data (age, gender, school grades, study path,...) a questionnaire was used in the second week of the semester.

Additionally to the covariates, the students finished five items of the Third International Mathematics and Science Study (TIMSS, Baumert et al.

(1999); Mullis et al. (2007)) in the questionnaire. The international scale of TIMSS is set to a mean of 500 with the standard deviation of 100 (Adams, Wu, & Macaskill, 1997). Because I only use a set of five items[1], which are suitable for the contents of the Analysis 1 lecture, I built sum scores of the correct answered items.

Possible predictors

With the study design, including the questionnaire, I take into account the following variables as students' attributes for the predictions. Different performance measures from school ('GPA', math grade in the final exam ('math grade'), average math grade in the last two years of school ('math av.')) and results of the TIMSS items ('timss'). The participants 'age' and 'sex'. The federal state ('state'), 'school type' and 'year' in which the university-entrance diploma was received. The variable 'school type' indicates if the the university-entrance diploma was received at a general-education Gymnasium (academic track). If the participants are teacher candidates (B.Ed.) or not ('tea'). If a prep course for math prior to the Analysis 1 lecture was attended ('prep') and if the respective semester was the first semester of a mathematical study program ('first'). The variable 'first' includes students which already attended other lectures than the Analysis 1 or already attended the Analysis 1 lecture but did not achieve exam admission and thus can not be recognized as former Analysis 1 participants.
The dependent variable 'pass' indicates the success in the Analysis 1 lecture. For a successful participation in the Analysis 1 the participants have to pass the final exam. The possible predictor variables are summarized in table 3.1.

[1] K4, K5, K6, L5, L6 of TIMSS/III

Table 3.1: Independent and dependent variables including description and scale

variable	description	scale
GPA	grade point average of school	1-6 with 1 as the best score
math grad	math grade of the final exam in school	1-15 with 15 as the best score
timss	sum score of the five TIMSS items	1-5
age	the participants age	
sex	the participants gender	m: male or f: female
school type	school type in which the the university-entrance diploma was received	1: general-education Gymnasium, 0: other
year	year in which the the university-entrance diploma was received	
math av.	average math grade in the last two years of school	1-15 with 15 as the best score
state	federal state in which the the university-entrance diploma was received	1: Baden-Württemberg, 0. other
prep	attendance of a mathematical prep course	1: yes, 0: no
tea	teacher candidate	1: yes, 0: no (B.Sc mathematics or physics)
first	first semester of a mathematical study program	1: yes, 0: no
pass	success in Analysis 1	1: pass, 0: fail

Redundant variables

I remove redundant variables with an absolute correlation of 0.75 or higher. According to the correlation matrix in table 3.2 one of the redundant variables 'year' or 'age' should be removed. Because of the better interpretability of the participant's age I remove the variable 'year'. Even though the absolute correlation value of 'math av.' with 'GPA' and 'math grade' is below 0.75 (0.71 and 0.74 respectively), I remove 'math av.' because of the high number of missing values. In table 3.3 the descriptive analysis of the remaining variables is outlined are reported for the different data sets.

Table 3.2: Variable correlation matrix of the combined dataset dataC

	GPA	school type	math grade	timss	age	sex	year	math av.	state	prep	tea	first
GPA	1.00											
school type	-0.17	1.00										
math grade	-0.67	0.21	1.00									
timss	-0.32	0.26	0.39	1.00								
age	0.13	-0.21	-0.13	-0.22	1.00							
sex	0.19	-0.08	-0.09	0.17	0.04	1.00						
year	-0.09	0.07	0.07	0.16	$\boxed{-0.94}$	-0.02	1.00					
math av.	-0.71	0.16	0.74	0.36	-0.06	-0.14	-0.00	1.00				
state	0.13	0.02	-0.06	0.03	-0.03	-0.09	0.02	-0.03	1.00			
prep	-0.06	-0.02	0.06	0.19	0.02	0.20	-0.03	0.09	-0.01	1.00		
tea	-0.03	0.07	-0.10	-0.20	0.03	-0.35	-0.01	0.01	0.09	-0.30	1.00	
first	0.06	-0.05	-0.01	0.06	-0.06	0.17	0.10	-0.06	-0.00	0.10	-0.28	1.00

Note. GPA: Grade point average; school type: general-education Gymnasium or not; math grade: math grade of the final exam in school; timss: sum scores of the TIMSS items; year: Year in which the final exam in school was done; math av.: Average math score of the last two years at school; state: school in Baden-Württemberg or not; prep: mathematical prep course or not; tea: B.Ed or not; first: first semester or not

3.2.3 Definition of dropout

In this study I use a very simple definition of dropouts. Students drop out if they fail the Analysis 1 lecture. The participants pass the Analysis 1 lecture if they qualify to take part at the final exam and pass the test. If they don't pass the final exam, they have one more chance in an repeat exam, which is similar to the original test. The admission to the exam depends, for example, on the grades on the homework students have to hand in weekly on the tutorials. Thus there are several ways to drop out. First, the students can choose to voluntary quit during the semester. Secondly, they might not obtain the admission for the final exam, or thirdly, they don't pass both the final exam and the repeat exam. In this study I don't differentiate between the different ways to drop out. I only consider the dichotomous variable pass or not pass.

3.2.4 General analysis

In this part I discuss general methods and procedures which occur in all the following models and algorithms. More details for the specific algorithms are discussed in the follow section. As explained bevor, the data set can be divided in cohort 1 (Analysis 1 lecture of the winter semester 2014/15),

Table 3.3: Descriptive data for cohort 1, cohort 2 and the combined data set dataC

	cohort 1	cohort 2	dataC
	M (SD)	M (SD)	M (SD)
GPA	2.09 (0.63)	1.95 (0.60)	2.02 (0.62)
math grade	11.56 (2.91)	11.94 (2.51)	11.76 (2.72)
timss	2.37 (1,23)	2.30 (1.19)	2.33 (1.21)
age	20.26 (2.47)	20.59 (5.53)	20.43 (4.27)
	% (n)	% (n)	% (n)
sex (male)	58.72 (101)	56.14 (96)	57.43 (197)
school type	76.74 (132)	78.95 (135)	77.84 (267)
state	86.05 (148)	87.72 (150)	86.88 (298)
prep	37.21 (64)	33.33 (57)	35.28 (121)
tea	43.60 (75)	51.46 (88)	47.52 (163)
first	67.44 (116)	77.19 (132)	72.30 (248)
pass	43.60 (75)	43.27 (74)	43.44 (149)
N	172	171	343

Note. GPA: Grade point average; math grade: math grade of the final exam in school; timss: sum scores of the TIMSS items; school type: general-education Gymnasium or not; state: school in Baden-Württemberg or not; prep: mathematical prep course or not; tea: B.Ed or not; first: first semester or not

cohort 2 (Analysis 1 lecture of the winter semester 2015/16) and the combined data set dataC. Due to the desired independence of lecturers and specific lecture schedules (e.g. the required achievements during the semester to achieve the admission for the final exam) the main data set is dataC. In order to report realistic measures for the prediction quality I divide the data set in a test set and a training set by randomly assigning 20% of the data to the test set. This procedure is performed with cohort 1 and cohort 2. Then the training sets and test sets are combined respectively to receive the training set and test set for dataC. Thereby I obtain the possibility to report measures on the test set (e.g. generalization error) of dataC separated in data derived from cohort 1 and cohort 2. The test set remains untouched and unseen until the evaluation of the specific algorithm. Table 3.4 illustrates the data splits.

Table 3.4: Configuration of the used data sets of cohort 1, cohort 2 and their combination dataC

	cohort 1	cohort 2	dataC
analysis 1 (2014/15)	X		X
analysis 1 (2015/16)		X	X
N (train)	138	137	275
N (test)	34	34	68

Differences in the prerequisites for teacher candidates

I compare the variable means and frequencies of the B.Ed. students with those of the B.Sc. students. Differences are testet with t-tests and χ^2-test.

Procedure for the prediction models

In the prediction models I train classifiers to predict the target[2] 'pass'. These are binary classifiers with the positive class referring to passing the Analysis 1 lecture and the negative class referring to dropout. The general procedure for the predictor models is as follows: I train the model using the training set of dataC. If hyperparameters need to be tuned I use cross-validation within the training set. For the model selection I use different prediction measures using the training set. In a second step, after the model selection, I report the prediction measures on the test set for model evaluation. For all methods I start with the biggest model using all available attributes as features.

$$\text{pass} \sim \text{GPA} + \text{school type} + \text{math grade} + \text{timss} + \text{age}$$
$$+\text{first} + \text{sex} + \text{state} + \text{prep} + \text{tea}$$

For feature selection I train the model with this selection, evaluate the prediction measures and check if the evaluation – the quality of the prediction – decreases for the smaller model. I apply this procedure to the dataset dataC and use the resulting selection for the rest of the analysis. The detailed feature selection procedure is discussed in the respective sections of the different

[2]In machine learning literature the dependent variable is often referred to as target (variable).

methods and algorithms.

Model evaluation – prediction measures

For the evaluation of the model I use different measures. I report the accuracy on the training set, the leave-one-out cross-classification/validation (loo.cv), and the precision, recall and F_1-score values. In the end I report the same measures, except for the loo.cv, on the separated test set.

While the accuracy measures the percentage of correctly classified outcomes, the precision gives true positives as a percentage of predicted positives and recall gives true positives as a percentage of actual positive outcomes. The F_1-score combines those measure to

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

For extremely skewed data the accuracy measure could be misleading[3] whereas the F_1-score can deal with skewed data. Even though the used data set is not extremely skewed, I report the F_1-score to control for possible effects. Apart from that, the F_1-score will play a minor role in the discussion.

The theoretically expected relation of the accuracy measures would be from high to low: the accuracy on the training set (as the parameters are trained for that set), the loo.cv accuracy (as it is measured on unseen data but within the dataset that was used for hyperparameter tuning) and then the accuracy on the test set (which is never seen before by the algorithm). Note that due to the relatively small size of the test set, I also consider the loo.cv as measure for the generalization error or accuracy. This error is know to be an unbiased estimator for the generalization performance of a classifier trained on m-1 examples (e.g. Evgeniou, Pontil, & Elisseeff, 2004; Rakotomamonjy, 2003). In Addition to the accuracy measures I report Cohen's Kappa as measure for the inter-rater agreement of the predictions and the true outcomes (Cohen, 1960) for both the training set and the test set. For

[3]e.g. an algorithm which simply assigns the positive class for every example would have the accuracy of 98% on an extremely skewed data set where only 2% of the training examples are in the negative class

the interpretation of Cohen's Kappa I use the suggestion of Landis and Koch (1977). They define the strength of agreement for kappa values of $0.00-0.20$ as slight, $0.21-0.40$ as fair, $0.41-0.60$ as moderate $0.61-0.80$ as substantial and $0.81 - 1.00$ as almost perfect.

Bounds for prediction accuracies

To gain a better sense of prediction accuracies I discuss some bounds in this section. As a lower bound of the expected prediction accuracy I could use a model that performs 50% guesses for every example. This method would result in an accuracy of .5 or 50%. But I use the accuracy of a baseline model which predicts a dropout for every example. With the 'pass' percentage of 43.44% (see table 3.3) in dataC the results of this model will have an accuracy of .57 or 56.56%. Therefor I can achieve an accuracy of 56.56% with a model that uses no information of the training data and thus builds the lower bound for the accuracies of our models.

The upper bound can not be specified exactly but I discuss some ideas. As features I only use attributes of the students bevor they came to university. Therefore this approach uses no information on the behavior of the students during the semester. But the active participation in lectures and tutorials as well as the general effort of the students is seen to be crucial for the success in math. Therefore the expected accuracies of our models are far less than 100%. Even with a sufficient amount of data, including data referring to the behavior during the semester, I would not expect to come close to 100%, because the final exam itself implies uncertainty of success. In conclusion I expect the accuracies to be better than 57% in order to have a valid predictor, but I do not expect the accuracies to exceed 80% (as an educated guess), due to the uncertainty of the behavior during the semester and in the test situation.

3.2.5 Analysis for the different methods

In this part I introduce the methods for the different algorithms and models. I use (i) the basic logistic regression, (ii) logistic regression with elastic-net regularization, (iii) the Support Vector Machine and (iv) tree based methods for feature selection and prediction. Survival analysis methods are applied solely for group difference and the effects of different features. Survival analysis methods (e.g. the Cox regression) predict the specific time of dropouts and therefore are not in the same accuracy frame as the other methods. Thus those methods are not used for prediction. Note that group differences and effects of features are also considered in the prediction models by the independent variables.

Logistic regression

Using logistic regressionI first include all features and report prediction measures. Then I repeat this approach only using features with either significant coefficients or at least features with almost significant coefficients. An analysis of deviance table is used to compare the respective model with all the features and with only the selected features. The prediction measures include results on the training set, as well as results on the test set. The test results are further differentiated in test examples from cohort 1 and cohort 2.

Logistic regression - elastic net regularization

The objective function for parameter estimation for penalized logistic regression uses the negative binomial log-likelihood

$$\min_{(\theta_0,\theta^{[-0]})\in\mathbb{R}^{n+1}} -\frac{1}{m}\left[\text{ log-likelihood }\right] + \lambda\left[(1-\alpha)||\theta||_2^2 + \alpha||\theta||_1\right]$$

where alpha controls for the tradeoff between L_2/ridge-regularization ($\alpha = 0$) and L_1/lasso-regularization ($\alpha = 1$) (Friedman, Hastie, & Tibshirani, 2010). The intercept θ_0 of the parameter vector $(\theta_0, \theta^{[-0]}) = \theta \in \mathbb{R}^{n+1}$ is not regularized.

I apply this method for different values of alpha ($\alpha \in \{0, 0.3, 0.6, 1\}$) and use the misclassification error as criterion for cross-validation to set the value for the regularization parameter λ. Due to the relatively small dataset I use leave-one-out cross validation (m-fold cross validation with the number of training examples m). Because of the properties of the L_1-norm the lasso performs a kind of continuous subset or feature selection (Hastie, Tibshirani, & Friedman, 2009). In the pure L_2 scenario ($\alpha = 0$) this feature selection part is lost. In that case I use the selected features of the basic logistic regression as comparison.

Survival analysis

In the survival analysis events are examined. The event in this analysis is a dropout. The 'time' variable – necessary in the survival analysis – refers to the time during the semester in weeks. A dropout at week x is defined as follows: to gain admission to the final exam the students have to gather points on the problem sets. If the admission to the final exam is not granted, the following week of the last submitted problem set is defined as the dropout time. If the admission is granted but the exam is failed, the last week of the semester indicates the dropout time.

I use the log rank test to evaluate whether the Kaplan-Meier survival curves for different groups are statistically equivalent (Kleinbaum & Klein, 2012). Additionally I use χ^2 tests of independence between features and the event (Terry M. Therneau & Patricia M. Grambsch, 2000; Therneau, 2015). The log rank test is based on the χ^2-test, comparing observed and expected cell counts through the categories of the outcome, here pass or dropout. In contrast to the χ^2 test of independence, which only includes total numbers at the end of the semester, the log-rank test includes the dropout times in the expected cell count. In both cases the null-hypothesis refers to independence, meaning that non significant results point to non significant influence of the feature on the occurrence of the event. Note that due to small cell sizes some features had to be grouped.

To obtain further measures of the influence of the features, I apply a Cox regression (Cox, 1972). In the Cox regression the Hazard function

$$h(t) = \frac{\text{number of persons with event in the interval beginning at } t}{(\text{number of surviving persons at } t) \cdot (\text{width of the interval})}$$

is used in the Cox model

$$h(t) = h_0(t) \cdot \exp\left(\sum_{i=1}^{n} \beta_i x_i\right) = h_0(t) \cdot \prod_{i=1}^{n} \exp(\beta_i x_i),$$

with the baseline-hazard $h_0(t)$. I use

$$\ln h(t) = \ln h_0(t) + \sum_{i=1}^{n} \beta_i x_i$$

with two groups A and B and the assumption for the hazard ratio:

$$HR = \frac{h_A(t)}{h_B(t)} = \text{constant}$$

Again, because of the prediction of dropouts as events in the survival analysis, the interpretation of the resulting coefficients has to be adapted. Negative coefficients point to a lower probability of the occurrence of the events, which means a higher probability of surviving in the survival analysis context or of success in the framework of this study. For example a coefficient of -0.5 for a grouping variable would refer to a Hazard-ratio of $\exp(-0.5) = 0.61$ which means the risk for the occurrence of the event is 39% lower for group 1 (compared to group 0).

Support Vector Machine (SVM)

In this section I use different SVMs to predict the outcome for the two Analysis 1 lectures. For the analysis I use the R package e1071 (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2017; R Core Team, 2015). I calculate both, a SVM with linear kernel and a SVM with radial basis function kernel (RBF kernel). In both cases I use 10-fold cross-validation on the training set for hyperparameter tuning. First I perform hyperparameter tuning using all the features. Then the resulting hyperparameters are used for the feature selection algorithms. With the resulting best feature subset, hyperparameter tuning is repeated and results are reported.

In the following I introduce the basic ideas of SVMs. Due to the mathematical complexity of the topic, only an overview is given. More details can be found in Hastie et al. (2009), for example. The goal of the SVM is to find the separating hyperplane with the biggest possible margin between the training points of the two classes. If the data is linearly separable, the support vector criterion for m training examples is usually written as

$$\min_{\beta,\beta_0} ||\beta||$$

$$\text{s.t. } y_i(x_i^T\beta + \beta_0) \geq 1, \forall i \in \{1, ..., m\}$$

with the separating hyperplane $\{x : f(x) = x^T\beta + \beta_0 = 0\}$, using the parameter vector $\beta \in \mathbb{R}^n$ and the intercept parameter $\beta_0 \in \mathbb{R}$, the feature vectors $x_i \in \mathbb{R}^n$ and the prediction outcome y_i (here $y_i \in \{-1, 1\}$). $G(x) = \text{sign } x^T\beta + \beta_0$ is used as classification rule.

With simple geometric considerations we can see that this fulfills the biggest possible margin notion. For example, if only considering the $y_i = 1$ case, the constraint can be rewritten to $(1, x_i)^T(\beta_0, \beta) \geq 1$ with $(1, x_i), (\beta_0, \beta) \in \mathbb{R}^{n+1}$. The inner product $(1, x_i)^T(\beta_0, \beta)$ can be written as $(1, x_i)^T(\beta_0, \beta) = p_i \cdot ||(\beta_0, \beta)||$ with the length of the projection p_i of the feature vector $(1, x_i)$ onto (β_0, β). Due to $(\beta_0, \beta)^T x = 0$ for all x on the hyperplane (meaning for all x with $f(x) = x^T\beta + \beta_0 = 0$), p_i gives the orthogonal distance of the feature vector $(1, x_i)$ to the hyperplane. With this context, the minimization of $||\beta||$ by simultaneously fulfilling the constraint results in maximizing the the distances to the hyperplane, thus maximizing the margin.

In the more common case of non perfectly linear separable data (in the features space) one still tries to maximize the margin, but has to allow some points in feature space to be on the wrong side. In this case the stack variables $\xi = (\xi_1, ..., \xi_m)$ are defined and the criterion is rewritten to

$$\min ||\beta||$$

$$\text{s.t. } y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i, \forall i \in \{1, ..., m\}$$

$$\xi_i \geq 0 \ \sum_i \xi_i \leq \text{constant}$$

This is quadratic with linear inequality constraints and thus results in a convex optimization problem which can equally written as

$$\min_{\beta,\beta_0} \frac{1}{2}||\beta||^2 + C\sum_{i=1}^{m}\xi_i$$

$$\text{s.t. } \xi_i \geq 0,\ y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i,\ \forall i$$

where the cost parameter C is introduced (which replaces the constant in the upper formulation). In the notion of Lagrange multipliers this can be written as the Lagrange function

$$L = \frac{1}{2}||\beta||^2 + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i\left[y_i(x_i^T\beta + \beta_0) - (1 - \xi_i)\right] - \sum_{i=1}^{m}\mu_i\xi_i$$

with the Langrange multipliers α_i for the inequality constraints and μ_i for the equality constraints. L will be minimized with respect to β, β_0 and ξ_i. In the following I will use the dual Lagrange objective function

$$L_{dual} = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j < x_i, x_j >$$

which now will be maximized subject to $0 \leq \alpha_i \leq C$ and $\sum_{i=1}^{m}\alpha_i y_i = 0$. Note that this representation only uses the inner product $< x_i, x_j >$ of the feature vectors instead of the single feature vectors x_i, which is an important property for kernel methods. For better readability I substitute some of those terms. I introduce the kernel function $K(x_i, x_j)$ (a positive (semi-) definite function) instead of $< x_i, x_j >$, $Q_{ij} = y_i y_j K(x_i, x_j)$ and the unity vector e. With the notion of the minimization of a cost function, I also use $J = -J_{dual}$ to obtain

$$\min_{\alpha} J = \min_{\alpha} \frac{1}{2}\alpha^T Q\alpha - e^T\alpha$$

$$\text{s.t. } 0 \leq \alpha_k \leq C \text{ and } y^T\alpha = 0$$

As mentioned earlier I use a linear kernel and the radial basic kernel which concerns the introduced kernel function. So far the support vector classifier finds linear boundaries in the feature space. For more flexibility the feature space can be enlarged by basic expansions (e.g. polynomials), which lead to general linear boundaries in the enlarged feature space and translate to nonlinear boundaries in the original feature space. In order to do so we can define basic functions $h_j(x)$ $(j = 1, ...M)$ and use the exact same procedure as described earlier with the substitution of the original features x_i to $h(x_i) = (h_1(x_i), ..., h_M(x_i))$ for all $i = 1, ..., m$. The SVM extends this idea by allowing arbitrary large dimensions for the enlarged feature space. As seen earlier the algorithm can be written entirely in terms of the kernen function K, meaning that only the kernel function $K(x_i, x_j) =< h(x_i), h(x_j) >$ has to be specified instead of the transformations $h(x_i)$.

SVM - linear kernel and SVM-RFE

In the linear case I use the kernel $K(x_i, x_j) = x_i^T x_j$ (which can be referred to as no kernel). The only hyperparameter is the cost parameter C. I apply cross-validation for a range of $C = 0.1, 0.2, ..., 0.9, 1, 2, 3, ..., 100$.

For feature selection I implement the recursive feature elimination algorithm SVM-RFE, described in Guyon, Weston, Barnhill, and Vapnik (2002), in combination with the e1071 package. In each iteration this algorithm ranks the features according to their influence on the weights and eliminates the last ranked feature until no feature is left. Then each subset is trained on the training set and evaluated using the loo.cv.

SVM - RBF kernel and feature selection for kernel methods

In the case of the RBF kernel I use

$$K(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right) = \exp\left(-\gamma||x_i - x_j||^2\right),$$

which means the cost parameter C and the shape parameter γ of the kernel have to be specified. In order to find the best hyperparameters I follow Meyer et al. (2017) and perform a 10-fold cross-validation using a grid search for the ranges $\gamma = 10^{-10}, 10^{-9}, ..., 10^2$ and $C = 0.1, 0.2, ..., 0.9, 1, 2, 3, ..., 100$, instead of first choosing a range for C and then choose γ for a preselected range of the C parameters (Chang & Lin, 2011).

To use SVM-RFE in the non-linear case, the approach can be generalized following Guyon et al. (2002). It is suggested (Kohavi & John, 1997) to use the change in the objective function, the cost function J, when one feature is removed as a ranking criterion. The authors of the Optimal Brain Damage algorithm (OBD algorithm, LeCun, Denker, & Solla, 1990), which approximated $DJ(i)$ by expanding J in Taylor series to second order, suggest to use $DJ(i)$ instead of the magnitude of the weights. The first order of the Taylor series can be neglected at the optimum and one gets

$$DJ(i) = \frac{1}{2}\frac{\partial^2 J}{\partial \beta_i^2}(D\beta_i)^2,$$

with the weight of the ith feature β_i. For the linear SVMs this is equivalent to using $(\beta_i)^2$ as ranking criterion (Guyon et al., 2002). This method can be extended to the non-linear case and to all kernel methods. One can assume no change in the value of the α's (the Lagrange multipliers for the inequality constraints), which makes the computations cheaper.

In the case of SVMs we obtain the cost function

$$J = \frac{1}{2}\alpha^T Q\alpha - e^T\alpha$$

$$\text{s.t. } 0 \leq \alpha_k \leq C \text{ and } y^T\alpha = 0$$

where e is the unity vector, C is the upper bound, $Q_{ij} = y_i y_j K(x_i, x_j)$ is a matrix and K is a kernel function. Recall, here I use the RBF kernel

$$K(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right) = \exp\left(-\gamma||x_i - x_j||^2\right)$$

with the shape parameter of the Gaussian kernel σ, or the kernel parameter γ like in e1071. In the linear SVM we had $K(x_i, x_j) = x_i^T x_j$ and $\alpha^T Q\alpha = \|\omega\|^2$, hence $DJ(i) = \frac{1}{2}(\omega_i)^2$.

I compute the change in cost function, caused by removing input i, by leaving the α's unchanged and re-compute the matrix Q. That means computing $K(x_h^{(-i)}, x_k^{(-i)})$ to get $Q^{(-i)}$, where $x^{(-i)}$ means that component i has been removed. As ranking coefficient we get:

$$|DJ(i)| = \left|\frac{1}{2}\alpha^T Q\alpha - \frac{1}{2}\alpha^T Q^{(-i)}\alpha\right|$$

Then I remove the input corresponding to the smallest difference $|DJ(i)|$. I iterate this procedure to carry out the RFE. If F denotes the set of all available features, we end up with subsets

$$F_1 \subset F_2 \subset ... \subset F_{n-1} \subset F$$

where F_i consists of i features. Like in the linear kernel case, for every subset F_i I train the algorithm and evaluate it with the loo.cv.

Tree based models

In this section I apply tree based models. I briefly review the basic concepts of those models and refer to Hastie et al. (2009) for a detailed, general introduction and to Hothorn, Hornik, and Zeileis (2006) and Hothorn, Buehlmann, Dudoit, Molinaro, and Van Der Laan (2006) for an introduction of the here used conditional inference trees and forests. Tree based methods in general – for example a single decision tree – partition the feature space into rectangles and fit simple models on each of them. In recursive binary partitioning the feature space is first split into two regions, where the feature and the split-point are selected by some fit measure. Then, the resulting regions could be split again using the same procedure. This is repeated until a stopping rule is applied. This results in the partitioning of the feature space into terminal nodes (the final branches after the respective last split). Each training example – for example participants in the Analysis 1 lecture – can be assigned (via the location of their feature vector in feature space) to one of the terminal nodes, where the simple classification or regression rule for this subsample is applied. The idea of ensemble tree methods – here random forests – is to overcome high variance in the single decision trees by growing a lot of trees and by averaging (in regression) or voting (in classification) to gain the overall output. In order to obtain a variety of different single trees, the random forest approach uses randomly selected feature subsets in each split as a basis for the splitting criteria.

I first evaluate a single tree. To avoid the well known problems of overfitting and selection bias towards features with many possible splits, as described in 'CART' (Breiman, Friedman, Olshen, & Stone, 1984) and 'C4.5' (Quinlan, 1993), I use a conditional inference tree implemented in the R package party (Hothorn, Hornik, & Zeileis, 2006; R Core Team, 2015). Secondly, I apply the random forest implementation based on the conditional inference trees (Hothorn, Buehlmann, et al., 2006; Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008; Strobl, Boulesteix, Zeileis, & Hothorn, 2007). A cross-validation approach is used to set the value for the number of features randomly selected in each split (mtry). I use the leave-one-out error to determine the best performance. For the other hyperparameters I use the suggested parameters, set as default in the party package (Hothorn, Buehlmann, et al., 2006; Strobl et al., 2008, 2007).

Conditional inference tree The feature selection in a single decision tree is included in the general concept as the features selected for the splits in the feature space partitioning. I report the selected features of the conditional inference tree which are used for the splits under the default hyperparameters.

Conditional forest For feature or variable importance in the conditional forests I use the conditional importance based on the permutation-importance measure, as described in (Strobl, Hothorn, & Zeileis, 2009). Due to correlating features I use the conditional importance instead of the original permutation-importance measure. Different random seeds are checked to ensure stability at least in the top ranked features. I use this ranking to measure accuracies on smaller subsets of the features.

3.2.6 Comparison of the predictors and risk groups

In the last step of the analysis I select the best predictors of each method, summarize their results and use them combined as ensemble predictor. For that purpose, I use the predictions of the single predictors and combine them via a majority vote for the class assignment.

I use this ensemble predictor for the identification of risk groups.

With the predictions the data is separated naturally into three groups. The first group consists of to the true positive predictions, including feature vectors of students for which the success is predicted and matches the true outcome. The second group contains the true negative predictions, including students which failed the Analysis 1 lecture and for which the failure was predicted. At last, the third group refers to students with feature vectors that were predicted wrong. Depending on the application, either group two or both group two and three can be defined as the risk group. In this study I define group two – the true negative prediction – as the risk group. Reasons for this choice are given by the general design of the prediction task. I don't include important information about students' behavior and situation during the semester, but only use information prior to university, thus wrong predictions can be partly associated with the information not included in this study. Therefore group three – the wrong predictions – refer to the uncertainty during the semester. In that context and at this point I can define the risk group as those students that are predicted to fail the lectures by the prediction model with high certainty.

I summarize descriptive measures for the three groups in the results. In order to correctly interpret the the results, it is important to define characteristics of the risk group. This can be done by partitioning the feature space and identifying the partitions, which can then be assigned to the risk group. Since this is exactly what the decision tree method does, I apply the already introduced conditional inference tree. For that purpose, I only include features that have been selected as most predictive by the different methods.

3.3 Results

In this part I first report results of the group comparison between B.Ed. and B.Sc. students. Secondly, I report the results of the different methods and thirdly, I discuss the results in total and compare the different methods I apply.

3.3.1 Prerequisites of teachers

Differences in prerequisites of teachers are shown in table 3.5.

Table 3.5: Group differences between B.Ed. students (teachers) and B.Sc. students.

	B.Ed (teacher)	B.Sc			
	M (SD)	M (SD)	t	df	Pr(> |t|)
GPA[1]	2.02 (0.60)	2.02 (0.63)	0.04	341	0.97
math grade[1]	**11.43 (2.63)**	**12.06 (2.77)**	**2.14**	**341**	**0.03**
timss[1]	**2.07 (1.19)**	**2.57 (1.18)**	**3.85**	**341**	**0.00**
age	20.49 (2.91)	20.37 (2.21)	-0.27	341	0.79
	% (n)	% (n)	χ^2	df	Pr(> χ^2)
sex:male	**41.72 (68)**	**71.67 (129)**	**30.17**	**1**	**0.00**
school type:1	81.60 (133)	74.44 (134)	2.14	1	0.14
prep:1	**20.25 (33)**	**48.89 (88)**	**29.50**	**1**	**0.00**
first:1	**57.67 (94)**	**85.56 (154)**	**31.84**	**1**	**0.00**
pass	46.01 (75)	41.11 (74)	0.65	1	0.42
N	163	180			

Note. **bold**: p < .5;
GPA = grade point average in the final exam in school; math grade = math rad in the final exam in school; timss = sum score of the TIMSS items; age = participant's age; sex = participant's sex; school type = general-education Gymnasium or not; prep = math prep course prior to university or not; first = first mathematical semester or not; pass = successful participation in the respective Analysis 1 lecture
[1]due to the sex distribution we repeated the analysis controlling for sex with no changes for the significant statements

The non significant differences in the GPA confirm the results of Klusmann et al. (2009), but we see differences in the selected TIMSS items and the math grade. Those results don't necessarily contradict the results of Klusmann et al. (2009), since my results solely concern the students of mathematics. There are also significant differences in prep and first. 49% of the B.Sc. students attended a mathematical prep course compared to only 20% of the teacher candidates, while at the same time significantly more B.Sc. students are in their first mathematical semester.

3.3.2 Logistic Regression

Results of the logistic regression can be found in table 3.6 for all features.

Table 3.6: Logistic regression with dataC and all features (dataC complete).

	Estimate	Std. Error	exp(Estimate)	z value	Pr(>|z|)
(Intercept)	-2.2441	1.8873	0.1060	-1.19	0.2344
GPA	**-1.0341**	**0.3534**	**0.3555**	**-2.93**	**0.0034**
school type1	0.7268	0.4026	2.0684	1.81	0.0710
math grade	**0.1809**	**0.0798**	**1.1983**	**2.27**	**0.0234**
timss	**0.4639**	**0.1455**	**1.5903**	**3.19**	**0.0014**
age	-0.0007	0.0501	0.9993	-0.01	0.9895
first1	**-0.7766**	**0.3506**	**0.4500**	**-2.21**	**0.0268**
sex	-0.0936	0.3255	0.9107	-0.29	0.7738
state1	0.2666	0.4535	1.3055	0.59	0.5567
prep1	0.5001	0.3315	1.6488	1.51	0.1315
tea1	0.6614	0.3412	1.9374	1.94	0.0526

Note. **bold**: p < .5;
GPA = grade point average in the final exam in school; school type = general-education Gymnasium or not; math grade = math grade in the final exam in school; times = sum score of the times items; age = participant's age; first = first mathematical semester or not; sex = participant's sex; state = federal state of Baden-Württemberg or not; prep = math prep course prior to university or not; tea = teacher candidate or not

In addition to the features with significant coefficients – GPA, math grade, timss and first – I select the features school type and tea for the next model, as their respective p-values (0.07 and 0.05) are close to being significant.

In table 3.7 the results of the logistic regression for the selected features is presented. Note that in both analysis the GPA is coded from 1 to 6 with 1 being the best grade. This means that a better GPA by one grade results in 65% (or 63% for the model with the selected features) better odds for the success in the lecture. We can see that the same features remain significant.

Table 3.7: Logistic regression with dataC and selected features (dataC select).

	Estimate	Std. Error	exp(Estimate)	z value	Pr(>|z|)
(Intercept)	-2.0194	1.5004	0.1327	-1.35	0.1783
GPA	**-0.9832**	**0.3379**	**0.3741**	**-2.91**	**0.0036**
school type1	0.7018	0.3894	2.0174	1.80	0.0715
math grade	**0.1798**	**0.0789**	**1.1970**	**2.28**	**0.0227**
timss	**0.5005**	**0.1395**	**1.6496**	**3.59**	**0.0003**
first1	**-0.7516**	**0.3402**	**0.4716**	**-2.21**	**0.0272**
tea1	0.5843	0.3248	1.7938	1.80	0.0720

Note. **bold**: $p < .5$
GPA = grade point average in the final exam in school; school type = general-education Gymnasium or not; math grade = math grade in the final exam in school; times = sum score of the times items; first = first mathematical semester or not; tea = teacher candidate or not

The analysis of deviance for these two models (table 3.8) shows no significant difference, meaning we can use the sparse model.

Table 3.8: Analysis of deviance for the two feature sets complete and selected.

	Resid. Df	Resid. Dev	Df	Deviance	Pr($> \chi^2$)
complete feature set	264	284.49			
selected feature set	268	287.18	-4	-2.70	0.6097

I use both models to predict the target variable ‘pass’ on the training set and on the test set. The different prediction measures for these models are reported in table 3.9. Prediction measures on the test sets are further differentiated in test examples of cohort 1 and cohort 2. In both cases the loo.cv is around 0.75 with a slightly higher value for the smaller model with 0.76, however the difference might be to small to allow a interpretation. The generalization on the test sets also shows only slight differences, probably with small advantages for the complete model. Both models show a rather high variance (overfitting) with a decline of accuracy from about 0.75 to 0.68. Note the differences in the divided test measures. We can see that cohort 2 might be more structured resulting in a lower generalization error. The overfitting problem can also be seen in the κ-values, which drop from moderate agreement on the training sets (0.51 and 0.52) to fair agreement (0.37 and 0.34) on the test set.

Table 3.9: Prediction measures on the training set and on the test set for the two logistic regression models. Test set measures are further differentiated in cohort 1 and cohort 2.

	dataC complete	cohort 1	cohort 2	dataC select	cohort 1	cohort 2
acc.train	0.76			0.77		
kappa.train	0.51			0.52		
loo.cv	0.75			0.76		
P.train	0.73			0.74		
R.train	0.71			0.71		
F1.train	0.72			0.72		
acc.test	0.69	0.62	0.76	0.68	0.62	0.74
P.test	0.66	0.53	0.79	0.63	0.53	0.73
R.test	0.63	0.57	0.69	0.63	0.57	0.69
F1.test	0.64	0.55	0.73	0.63	0.55	0.71
kappa.test	0.37	0.22	0.52	0.34	0.22	0.47

Note. acc = accuracy; kappa = inter-rater agreement (Cohen's kappa); loo.cv = leave-one-out accuracy; P = precision; R = recall; F1 = F1 score

3.3.3 Logistic Regression - elastic net

For the logistic regression with elastic net regularization I use, as mentioned already, one model with pure lasso / L_1 ($\alpha = 1$) regularization, two mixed models with $\alpha = 0.6$ and $\alpha = 0.3$ and one model with pure ridge / L_2 ($\alpha = 0$) regularization. For the case of $\alpha = 0$, one more model is included using the selected features of the basic logistic regression, because we can not profit from the feature selection property of the L_1-regularization. Prediction results are presented in table 3.10, including result on the test set divided in the two cohorts and the used feature subsets. The respective selected subsets are presented in table 3.11 for different values of α.

Table 3.10: Prediction measures of the elastic net for dataC using different values of the tradeoff parameter α. Test set measures are further differentiated in cohort 1 and cohort 2.

	pure L_1			pure L_2	
	$\alpha = 1$	$\alpha = 0.6$	$\alpha = 0.3$	$\alpha = 0$	$\alpha = 0$
	$\lambda = 0.089$	$\lambda = 0.135$	$\lambda = 0.107$	$\lambda = 0.306$	$\lambda = 0.369$
acc.train	0.76	0.75	0.75	0.77	0.75
kappa.train	0.49	0.48	0.48	0.53	0.49
loo.cv	0.73	0.74	0.73	0.74	0.74
P.train	0.77	0.76	0.72	0.76	0.74
R.train	0.62	0.61	0.66	0.69	0.66
F1.train	0.69	0.68	0.69	0.72	0.70
acc.test	0.69	0.71	0.74	0.72	0.71
P.test	0.66	0.68	0.69	0.69	0.68
R.test	0.63	0.63	0.73	0.67	0.63
F1.test	0.64	0.66	0.71	0.68	0.66
kappa.test	0.37	0.40	0.47	0.43	0.40
cohort 1					
acc.test	0.59	0.62	0.71	0.65	0.68
P.test	0.50	0.55	0.62	0.57	0.62
R.test	0.43	0.43	0.71	0.57	0.57
F1.test	0.46	0.48	0.67	0.57	0.59
kappa.test	0.13	0.18	0.41	0.27	0.32
cohort 2					
acc.test	0.79	0.79	0.76	0.79	0.74
P.test	0.76	0.76	0.75	0.80	0.73
R.test	0.81	0.81	0.75	0.75	0.69
F1.test	0.79	0.79	0.75	0.77	0.71
kappa.test	0.59	0.59	0.53	0.59	0.47

Note. acc = accuracy; kappa = inter-rater agreement (Cohen's kappa); loo.cv = leave-one-out accuracy; P = precision; R = recall; F1 = F1 score; λ = regularization parameter

All models show comparable and rather good results on the training sets with accuracies between 0.75 and 0.77 and moderate agreement with the true output (κ-value). The loo.cv as well does not clearly prefer one of the models. Differences occur for the generalization on the test set, which was expected as we use different regularizations to prevent overfitting in all the models. The best results are shown by the model with $\alpha = 0.3$ on all prediction measures. Note that, except for the pure L_1 model ($\alpha = 1$), all inter-rater measures are in, or very close to the moderate agreement on the test set. For the division of the test results into the two cohorts, again the $\alpha = 0.3$ model shows the best results, meaning that even the cohort two, which seems to be harder to predict, is predicted with an accuracy of 0.71.

Table 3.11: Respective feature selection of the elastic net for dataC using different values of the tradeoff parameter α.

	pure L_1			pure L_2	
	$\alpha = 1$	$\alpha = 0.6$	$\alpha = 0.3$	$\alpha = 0$	$\alpha = 0$ *
GPA	X	X	X	X	X
school type	.	.	X	X	X
math grade	X	X	X	X	X
timss	X	X	X	X	X
age	.	.	.	X	.
first	.	.	X	X	X
sex	.	.	.	X	.
state	.	.	.	X	.
prep	.	.	.	X	.
tea	.	.	X	X	X

Note. * feature subset pre-design with results of the basic logistic regression
GPA = grade point average in the final exam in school; school type = general-education Gymnasium or not; math grade = math grade in the final exam in school; timss = sum score of the TIMSS items; age = participant's age; first = first mathematical semester or not; sex = participant's sex; state = federal state of Baden-Württemberg or not; prep = math prep course prior to university or not; tea = teacher candidate or not

The feature selection in models with $\alpha > 0$ is done by the algorithm as a property of the L_1-norm. The more weight is on the L_1-norm, compared to the L_2-norm (regulated by α), the more features are removed. Note that for the best model with $\alpha = 0.3$ exactly the same features are selected as in the approach with the basic logistic regression. But here, the models with this feature selection shows better generalization results than the basic logistic regression, because in both cases ($\alpha = 0.3$ and $\alpha = 0$) regularization is used.

3.3.4 Survival Analysis

The data used for the survival analysis (received from dataC) is represented as the survival in table 3.12.

Table 3.12: Survival table of dataC

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0	343	2	0.99	0.00	0.99	1.00
2	341	5	0.98	0.01	0.96	0.99
3	336	6	0.96	0.01	0.94	0.98
4	330	10	0.93	0.01	0.91	0.96
5	320	8	0.91	0.02	0.88	0.94
6	312	9	0.88	0.02	0.85	0.92
7	303	2	0.88	0.02	0.84	0.91
8	301	14	0.84	0.02	0.80	0.88
9	287	6	0.82	0.02	0.78	0.86
10	281	14	0.78	0.02	0.74	0.82
11	267	6	0.76	0.02	0.72	0.81
12	261	12	0.73	0.02	0.68	0.77
13	249	16	0.68	0.03	0.63	0.73
14	233	4	0.67	0.03	0.62	0.72
15	229	3	0.66	0.03	0.61	0.71
17	226	14	0.62	0.03	0.57	0.67
18	212	13	0.58	0.03	0.53	0.63
19	199	50	0.43	0.03	0.38	0.49

Note. time in weeks during the semester; n.risk = number of participants still in the lecture at that time; n.event = number of dropouts at that time; survival = percentage of participants still in the lecture

In this table ‘time’ refers to the weeks during the semester, ‘n.risk’ is the number of students in each week, which still participate in the lecture. The variable ‘n.event’ gives the number of events in this week. In the context of this study, this means the number of dropouts in this weeks. The percentage of participants still in the lecture is given by ‘survival’. Note that, due to the study design there is no censored data within the time interval of the semester. That means, the only participants with censored data are those with no event – no dropout – in the time interval, in this context those participants which pass the lecture. For the survival analysis the variables 'time' and 'event' are analyzed.

The independence of the different features of the target variable 'pass' or the time of the dropout is tested with the χ^2 test of independence and the log rank test. The results of the tests are shown in table 3.13.

Table 3.13: Results for independence tests of the variables and the event (χ^2) and group differences in the survival analysis (log rank)

	log rank test			χ^2 test		
	χ^2	df	p	χ^2	df	p
cohort	3.8	1	.051	0	1	1
sex	3.4	1	.066	1.79	1	.180
tea	1.3	1	.247	0.69	1	.421
state	0	1	.890	0.00	1	.988
prep	**4.3**	**1**	**.039**	2.5	1	.114
age[1]	1	3	.81	0.61	3	0.90
school type	**24.7**	**1**	**.000**	**18.76**	**1**	**.000**
first	**10.1**	**1**	**.002**	**8.87**	**1**	**.003**
timss	**37**	**5**	**.000**	**34.19**	**5**	**.000**
GPA[2]	**101**	**26**	**.000**	**55.42**	**4**	**.000**
math grade[3]	**69.5**	**11**	**.000**	**69.20**	**6**	**.000**

Note. **bold**: p < .5; [1] only age 18-21; [2] combined grades for χ^2: 1-1.4, 1.5-1.9, 2-2.4, 2.5-2.9, 3-3.6; [3] combined points for χ^2: <10, 11, 12, 13, 14, 15
cohort = cohort 1 or cohort 2; sex = participant's sex; tea = teacher candidate or not; state = federal state of Baden-Württemberg or not; prep = math prep course prior to university or not; age = participant's age; school type = general-education Gymnasium or not; first = first mathematical semester or not; timss = sum score of the TIMSS items; GPA = grade point average in the final exam in school; math grade = math grade in the final exam in school

In both tests the features 'school type', 'first', 'timss', 'GPA' and 'math grades' show significant dependence with the target variable 'pass'. The feature 'prep' shows a significant difference in the Kaplan-Meier survival curves (for 'prep'=0 and 'prep'=1), but no significant dependence to the target variable 'pass' using the χ^2 test. This could mean that the dropout times differ for those groups, but not the total number of dropouts at the end of the time interval of the semester.

The results of the Cox regression are shown in table 3.14. Again, 'GPA', 'math grade', 'timss' and 'first' show significant coefficients. As in the independence tests the 'school type' determines the dropout. Like already mentioned, I did not use the Cox regression as a predictor because here the survival probability at a time point is the dependent variable in contrast to 'pass' in the other algorithms.

Table 3.14: Results of the Cox regression with all features.

	coef	exp(coef)	se(coef)	z	p
GPA	**0.35**	**1.43**	**0.16**	**2.28**	**0.02**
school type1	**-0.43**	**0.65**	**0.17**	**-2.60**	**0.01**
math grade	**-0.09**	**0.91**	**0.03**	**-2.61**	**0.01**
timss	**-0.16**	**0.85**	**0.07**	**-2.19**	**0.03**
age	0.01	1.01	0.02	0.58	0.56
first1	**0.56**	**1.75**	**0.19**	**3.02**	**0.00**
sex	0.09	1.09	0.16	0.55	0.58
state1	-0.02	0.98	0.21	-0.09	0.93
prep1	-0.31	0.73	0.17	-1.85	0.06
tea1	-0.16	0.85	0.17	-0.95	0.34

Note. **bold**: $p < .5$
GPA = grade point average in the final exam in school; school type = general-education Gymnasium or not; math grade = math grade in the final exam in school; timss = sum score of the TIMSS items; age = participant's age; first = first mathematical semester or not; sex = participant's sex; state = federal state of Baden-Württemberg or not; prep = math prep course prior to university or not; tea = teacher candidate or not

3.3.5 SVM with linear kernel

For the feature selection in the linear SVM I apply the RFE algorithm described on page 106. Table 3.15 shows the resulting feature ranking and the selected subset.

Each feature subset is tested using the leave-one-out classification on the training set. Note that the ranking only marks the feature that is removed for the next subset. This evaluation is done within each subset. Rankings within the selected subsets are meaningless, except for the last ranked feature, which is chosen to be removed in the next step. This means in the best selected subset the feature 'school type' is marked as the least important feature. There is no ranking within this subset for the remaining features. In addition to the 'school type' the best subset consists of the 'math grade',

Table 3.15: Linear SVM-RFE: Ranking and selected subsets for dataC.

ranking and **best set**										C	loo.cv	loo.cv complete
math grade	**GPA**	**timss**	**tea**	**first**	**school type**	prep	sex	state	age	0.1	0.7782	0.76

Note. **bold** and framed = best feature subset selected by the linear SVM-RFE algorithm; C = cost/regularization hyperparameter; loo.cv = leave-one-out classification on the best selected set; loo.cv complete = leave-one-out classification on all features
GPA = grade point average in the final exam in school; school type = general-education Gymnasium or not; math grade = math grade in the final exam in school; timss = sum score of the TIMSS items; age = participant's age; first = first mathematical semester or not; sex = participant's sex; state = federal state of Baden-Württemberg or not; prep = math prep course prior to university or not; tea = teacher candidate or not

'GPA', 'timss', 'tea' and 'first'. As a comparison the loo.cv for the complete model (which has not been selected) is reported as well.

I use the linear SVM with the selected feature subset and the complete feature set to predict the target 'pass'. Prediction measures are summarized in table 3.16. The measures indicate that the smaller model achieves almost the same results as the complete model and thus can be used here. Both models show moderate inter-rater agreement ($\kappa = 0.44$) on the test set. The loo.cv of cause is higher for the smaller model due to the results of the RFE algorithm.

3.3.6 SVM with RBF kernel

For the feature selection for the SVM with RBF kernel the adjusted RFE algorithm, described on page 106 is used. Results are shown in table 3.17.

The same statements about the feature ranking described in the previous section hold here. The hyperparamters C and γ are tuned using cross-validation on the training set and are then used throughout the selection algorithm. The selected best feature subset is identical to the one in the linear case. Included in the selected feature subset are the 'math grade', 'GPA', 'timss', 'first', 'tea', and 'school type' resulting in a loo.cv value of 0.77 compared to a loo.cv value of 0.76 for the complete model.

Again I use both models, taking the complete feature set and the selected subset into account as predictors. The results are reported in table 3.18.

The generalized results of the smaller model are slightly better than of the

Table 3.16: Evaluation of the linear SVM model trained on dataC with the selected features and on dataC with the complete feature set. Test set measures for the selected feature set are further differentiated in cohort 1 and cohort 2.

	dataC (C = 0.2)	cohort 1	cohort 2	dataC complete (C=0.1)
acc.train	0.77			0.77
kappa.train	0.53			0.54
loo.cv	0.78			0.76
precision.train	0.73			0.73
recall.train	0.74			0.76
F1.train	0.74			0.75
acc.test	0.72	0.68	0.76	0.72
kappa.test	0.44	0.35	0.53	0.44
precision.test	0.67	0.59	0.75	0.66
recall.test	0.73	0.71	0.75	0.77
F1.test	0.70	0.65	0.75	0.71

Note. acc = accuracy; kappa = inter-rater agreement (Cohen's kappa); loo.cv = leave-one-out accuracy; P = precision; R = recall; F1 = F1 score; C = cost/regularization hyperparameter

Table 3.17: Generalized SVM-RFE for RBF kernels: Ranking and selected subsets for dataC

ranking and **best set**										C	γ	loo.cv	loo.cv complete
math grade	**GPA**	**timss**	**first**	**tea**	**school type**	age	sex	prep	state	28	0.001	0.7709	0.7564

Note. **bold** and framed = best feature subset selected by the generalized SVM-RFE algorithm; C = cost/regularization hyperparameter; γ = RBF kernel hyperparamter; loo.cv = leave-one-out classification on the best selected set; loo.cv complete = leave-one-out classification on all features
GPA = grade point average in the final exam in school; school type = general-education Gymnasium or not; math grade = math grade in the final exam in school; timss = sum score of the TIMSS items; age = participant's age; first = first mathematical semester or not; sex = participant's sex; state = federal state of Baden-Württemberg or not; prep = math prep course prior to university or not; tea = teacher candidate or not

complete model, with both models showing moderate inter-rater agreement (0.50 and 0.47 respectively). The models show slightly more overfitting than the logistic regression model with elastic net regularization, but result in comparable results of accuracies around 0.75. As expected, the SVM with

Table 3.18: Evaluation of the SVM models with RBF kernel trained on dataC with the selected feature set and the complete feature set. Test set measures for the selected feature set are further differentiated in cohort 1 and cohort 2.

	dataC (C=28, γ=0.001)	cohort 1	cohort 2	dataC complete (C=28, γ=0.001)
acc.train	0.78			0.78
kappa.train	0.56			0.56
loo.cv	0.77			0.76
precision.train	0.74			0.73
recall.train	0.76			0.78
F1.train	0.75			0.76
acc.test	0.75	0.71	0.79	0.74
kappa.test	0.50	0.41	0.59	0.47
precision.test	0.70	0.63	0.76	0.68
recall.test	0.77	0.71	0.81	0.77
F1.test	0.73	0.67	0.89	0.72

Note. acc = accuracy; kappa = inter-rater agreement (Cohen's kappa); loo.cv = leave-one-out accuracy; P = precision; R = recall; F1 = F1 score; C = cost/regularization hyperparameter; γ = RBF kernel/shape hyperparameter

the RBF kernel outperforms the linear SVM.

3.3.7 Conditional inference tree

The selected features by the conditional inference tree are shown in table 3.19.

Table 3.19: Selected features of the conditional inference tree on dataC. Those features are used for splits in the decision tree.

	math grade	GPA	timss	school type	first	age	prep	state	tea	sex
dataC	X	X	X						X	

Note. math grade = math grade in the final exam in school; GPA = grade point average in the final exam in school; timss = sum score of the TIMSS items; school type = general-education Gymnasium or not; first = first mathematical semester or not; age = participant's age; prep = math prep course prior to university or not; state = federal state of Baden-Württemberg or not; tea = teacher candidate or not; sex = participant's sex

The algorithm uses the often selected features ‘math grade’, ‘GPA’ , ‘timss’ and ‘tea’ to do the splits. The specific splits are shown in figure 3.2.

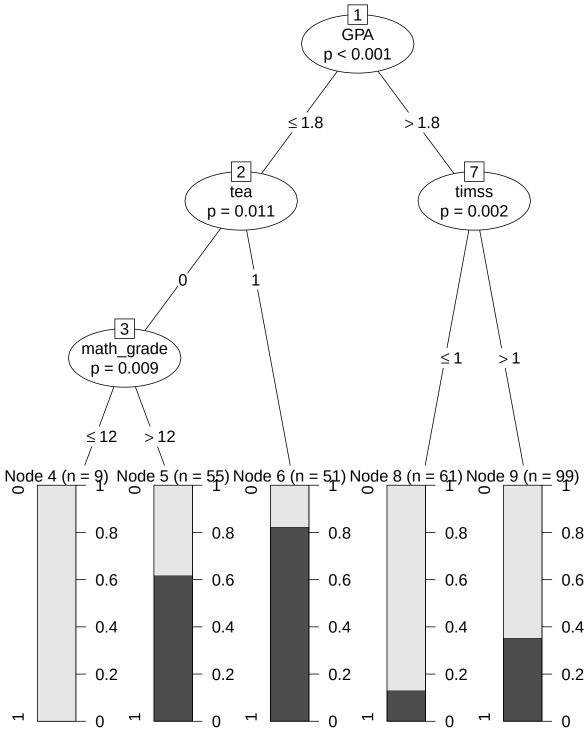

Figure 3.2: Conditional inference tree with the selected features and splits
Note. The figure shows the split-points within the features used for splits. In the terminal nodes 4-6, 8, 9 the frequencies of the target variable 'pass' is shown.
GPA = grade point average in the final exam in school; tea = teacher candidate or not; timss = sum score of the TIMSS items; math_grade = math grade in the final exam in school

The first split is done in node 1 for a GPA ≤ 1.8 (better GPA than 1.8), these participants are further divided in node 2 by their major (teacher or no teacher candidate). For the teacher candidates no further split is done and the success in the lecture is predicted. Students majoring in physics (B.Sc.) and math (B.Sc.), are clustered by the feature 'math grade' in node 3. Students with math grades of 12 points or lower form the failure class, students with higher math grads higher than 12 points the success class. Students with a GPA > 1.8 are sectioned once more in node 7 depending on wheater or not more than one TIMSS item was correctly answered. Note that even though node 7 executes a further split, when used as a predictor the decision tree predicts a failure for all participants in the two terminal nodes 8 and 9, because of pass frequencies lower than 0.5.

The prediction measures for the conditional inference tree are presented in table 3.20. The predictor shows overfitting, resulting in a total test accuracy of 0.68 and only a fair inter-rater agreement ($\kappa = 0.31$).

Table 3.20: Evaluation of the conditional inference tree as predictor. Test set measures are further differentiated in cohort 1 and cohort 2.

	dataC	cohort 1	cohort 2
acc.train	0.74		
kappa.train	0.47		
loo.cv	0.69		
precision.train	0.73		
recall.train	0.66		
F1.train	0.69		
acc.test	0.68	0.62	0.74
kappa.test	0.31	0.22	0.47
precision.test	0.64	0.53	0.71
recall.test	0.50	0.57	0.75
F1.test	0.56	0.55	0.73

Note. acc = accuracy; kappa = inter-rater agreement (Cohen's kappa); loo.cv = leave-one-out accuracy; P = precision; R = recall; F1 = F1 score

3.3.8 Conditional forests

The value of the number of features randomly selected in each split (mtry) is set via cross-validation to mtry= 2 (default value is mtry= 5). Feature selection depends on variable importance. Here I can see a slight dependence on the random seed. Those dependencies only occur within the second and third to last ranked features 'state' and 'prep'. Conditional variable importance is calculated three times with different random seeds. In table 3.21 the sorted, absolute values of the means are reported.

Even though there is no clear cut in feature importance, consistent to the other methods I select the ranked features 'school type' to 'math grade' and compare the model with these features to the complete model. I choose this

Table 3.21: Sorted list of the conditional variable importance for the conditional forest and the feature selection used in further analysis.

sex	prep	state	age	**school type**	**tea**	**first**	**timss**	**GPA**	**math grade**
0.00032	0.00083	0.00091	0.00127	0.00265	0.00418	0.00434	0.00733	0.01254	0.01802

Note. Absolute values of the means of three tested random seeds; bold and framed = selected subset for further investigations (dataC select)
GPA = grade point average in the final exam in school; school type = general-education Gymnasium or not; math grade = math grade in the final exam in school; timss = sum score of the TIMSS items; age = participant's age; first = first mathematical semester or not; sex = participant's sex; state = federal state of Baden-Württemberg or not; prep = math prep course prior to university or not; tea = teacher candidate or not

cut to compare the results with the SVM, where the same subset has been selected.

The prediction measures can be found in table 3.22. Even though the complete model shows better results on the training set, the results on the test set are slightly better for the smaller model.

Table 3.22: Accuracy measures of the random forest for all features (complete) und the selected most important features. Test measures for the complete model are further divided in the two cohorts.

	dataC complete	cohort 1	cohort 2	dataC select
acc.train	0.80			0.76
kappa.train	0.60			0.51
loo.cv	0.74			0.71
precision.train	0.78			0.73
recall.train	0.76			0.70
F1.train	0.77			0.72
acc.test	0.69	0.65	0.74	0.71
kappa.test	0.38	0.29	0.47	0.40
precision.test	0.65	0.56	0.73	0.67
recall.test	0.67	0.64	0.69	0.67
F1.test	0.66	0.6	0.71	0.67

Note. dataC select includes the features: 'school type', 'tea', 'first', 'timss', 'GPA' and 'math grade'
acc = accuracy; kappa = inter-rater agreement (Cohen's kappa); loo.cv = leave-one-out accuracy; P = precision; R = recall; F1 = F1 score

3.3.9 Comparison and ensemble of the best predictors

For each method I select the best model. For some methods with only slight differences on the prediction measures, I choose the smaller model. For the logistic regression I choose the model with the smaller subset selection. From the logistic regressions with elastic net regularization we choose the model with $\alpha = 0.3$. In both SVM algorithms – linear and RBF kernel –, as well as for the random forest method, I select the smaller model. Additionally the conditional inference tree is included. A summary of the prediction measures is reported in table 3.23.

Table 3.23: Summary of the prediction measures of the best predictors.

	logistic regression	log. reg. with el. net	linear SVM	SVM with RBF kernel	tree	forest
acc.train	0.77	0.75	0.77	0.78	0.74	0.76
kappa.train	0.52	0.48	0.53	0.56	0.47	0.51
loo.cv	0.76	0.73	0.78	0.77	0.69	0.71
P.train	0.74	0.72	0.73	0.74	0.73	0.73
R.train	0.71	0.66	0.74	0.76	0.66	0.70
F1.train	0.72	0.69	0.74	0.75	0.69	0.72
acc.test	0.68	0.74	0.72	0.75	0.68	0.71
kappa.test	0.34	0.47	0.44	0.50	0.31	0.40
P.test	0.63	0.69	0.67	0.70	0.64	0.67
R.test	0.63	0.73	0.73	0.77	0.50	0.67
F1.test	0.63	0.71	0.70	0.73	0.56	0.67

Note. logistic regression = basic logistic regression (selected subsets); log. with el. net = logistic regression with elastic net regularization ($\alpha = 0.3$); linear SVM = SVM with linear kernel (selected features; C=0.2); SVM with RBF kernel (selected features; C= 28, γ=0.001); tree = conditional inference tree; forest = random forest based on conditional inference trees (selected features)
acc = accuracy; kappa = inter-rater agreement (Cohen's kappa); loo.cv = leave-one-out accuracy; P = precision; R = recall; F1 = F1 score

Further information (like hyperparameters) about the single predictors can be found in the specific sections. Both SVM algorithms, the logistic regression with elastic net regularization and the random forest achieve test accuracies higher than 0.70 and moderate inter-rater agreement. The two outstanding algorithms are the logistic regression with elastic net regularization ($\alpha = 0.3$) and the SVM with RBF kernel. Both achieve the highest test accuracy (0.74 and 0.75 respectively) as well as the highest inter-rater agreement kappa.test (0.47 and 0.50 respectively).

To compare the predicted outcomes of the different models I calculate the inter-rater agreement between the models. Tables for those κ-values are presented in table 3.24 for Cohen's Kappa on the training set and in table 3.25 on the test set.

Table 3.24: Table of inter-rater agreement (κ) for the different best predictors on the training set

	log.train	elnet.3.train	svm.lin.train	svm.train	tree.train	forest.train
log.train	1					
elnet.3.train	0.88	1				
svm.lin.train	0.92	0.90	1			
svm.train	0.91	0.90	0.98	1		
tree.train	0.64	0.70	0.64	0.66	1	
forest.train	0.78	0.84	0.85	0.86	0.70	1

Note. log = basic logistic regression (selected subsets); elnet.3 = logistic regression with elastic net regularization (α = 0.3); svm.lin = SVM with linear kernel (selected features; C=0.2); svm = SVM with RBF kernel (selected features; C= 28, γ=0.001); tree = conditional inference tree; forest = random forest based on conditional inference trees (selected features)

On both, the training set and the test set, the different algorithms show at least substantial and often almost perfect inter-rater agreement.

Table 3.25: Table of inter-rater agreement (κ) for the different best predictors on the test set

	log.test	elnet.3.test	svm.lin.test	svm.test	tree.test	forest.test
log.test	1					
elnet.3.test	0.82	1				
svm.lin.test	0.85	0.97	1			
svm.test	0.79	0.97	0.94	1		
tree.test	0.70	0.76	0.74	0.79	1	
forest.test	0.68	0.85	0.82	0.88	0.91	1

Note. log = basic logistic regression (selected subsets); elnet.3 = logistic regression with elastic net regularization (α = 0.3); svm.lin = SVM with linear kernel (selected features; C=0.2); svm = SVM with RBF kernel (selected features; C= 28, γ=0.001); tree = conditional inference tree; forest = random forest based on conditional inference trees (selected features)

The best feature subsets for the methods are summarized in table 3.26.

Except for the single conditional inference tree, all of the most successful algorithms selected the same feature subset. This is remarkable, because

Table 3.26: Selected features for the different methods.

	logistic regression	log. reg. with el. net	linear SVM	SVM with RBF kernel	tree	forest
GPA	X	X	X	X	X	X
school type	X	X	X	X	.	X
math grade	X	X	X	X	X	X
timss	X	X	X	X	X	X
age	.	.	.	.	.	.
first	X	X	X	X	.	X
sex	.	.	.	.	.	.
state	.	.	.	.	.	.
prep	.	.	.	.	.	.
tea	X	X	X	X	X	X

Note. GPA = grade point average in the final exam in school; school type = general-education Gymnasium or not; math grade = math grade in the final exam in school; timss = sum score of the TIMSS items; age = participant's age; first = first mathematical semester or not; sex = participant's sex; state = federal state of Baden-Württemberg or not; prep = math prep course prior to university or not; tea = teacher candidate or not logistic regression = basic logistic regression; log. with el. net = logistic regression with elastic net regularization ($\alpha = 0.3$); linear SVM = SVM with linear kernel (C=0.2); SVM with RBF kernel (C= 28, γ=0.001); tree = conditional inference tree; forest = random forest based on conditional inference trees

the feature selection is done for each algorithm separately and with algorithm specific, appropriate methods. In addition to the performance measures ('GPA', 'math grade' and 'timss'), the school type ('school type'), first mathematical semester or not ('first') and the major ('tea') is selected. For the predictions, the major ('tea') is selected as feature, even though this is the only selected variable that shows no significant result, neither in the coefficients of the logistic regressions and the Cox regression, nor in the χ^2 and log rank tests in the survival analysis.

As a final result, I use all the predictors of table 3.23 to build an ensemble predictor. In the ensemble a majority vote is used to gain the overall prediction. In table 3.27 the prediction measures for this ensemble are reported.

The ensemble predictor does not outperform the best single predictors. This is no surprise, due to the high inter-rater agreements shown in table 3.24 and table 3.25. It more or less adopts the prediction measures of the SVM with RBF kernel and the logistic regression with elastic net regularization ($\alpha = 0.3$).

Table 3.27: Prediction measures of the ensemble of the best predictors. The ensemble consists of the predictors summarized in table 3.23. Prediction by majority vote of the single predictors.

	ensemble
acc.train	0.79
kappa.train	0.56
P.train	0.75
R.train	0.76
F1.train	0.76
acc.test	0.75
kappa.test	0.50
P.test	0.70
R.test	0.77
F1.test	0.73

Note. acc = accuracy; kappa = inter-rater agreement (Cohen's kappa); loo.cv = leave-one-out accuracy; P = precision; R = recall; F1 = F1 score

3.3.10 Identification and description of the risk group

For the identification of the different prediction groups (true positive predictions, true negative predictions and wrong predictions), I use the ensemble predictor to gain predictions for the complete dataset dataC. This dataset contains the training set and the test set, which results in accuracy measures different those reported in table 3.27 (closer to the training measures due to the distribution in the train-test split). The prediction measures for the combined data set (train set and test set) are reported in table 3.28.

In table 3.29 descriptive measures of the three groups are given. The performance measures from school – GPA and math grade –, as well as the test performance on the TIMSS items show the relation of performance and success. The group of the true predictions shows the best grades and test performances, whereas the risk group of the true negative predictions shows the weakest performance. The uncertainty group of the wrong predictions shows intermediate performance between the other two groups. With regard to the ranges of the performance measures, we can see that for students with a GPA worse than 2.4 and a math grade worse than 9 points the success is not

Table 3.28: Prediction measures of the ensemble predictor. The predictor was learned on the training set of dataC (for results see table 3.27). The here reported measures are on the complete dataset.

	complete dataC
acc	0.78
kappa	0.55
P	0.74
R	0.77
F1	0.75

Note. acc = accuracy; kappa = inter-rater agreement (Cohen's kappa); loo.cv = leave-one-out accuracy; P = precision; R = recall; F1 = F1 score

once correctly predicted. For students with a GPA better than 1.3 no correct failure prediction occurs. The ranges of the wrong predictions are rather wide. This could stress that importance of the behavior during the semester, as fortunate prerequisites don't necessarily lead to success and unfortunate prerequisites don't necessarily lead to failure. The descriptive measures in table 3.29 only give general information on the groups but no indication on the structure or interaction of the features, which lead to different group assignments.

Table 3.29: Descriptive measures of the groups with correct positive prediction, correct negative prediction and wrong predictions on the complete dataset dataC. The predictions are executed with the ensemble predictor.

	pred. pass=1		wrong pred.		pred. pass=0	
	M (SD)	range	M (SD)	range	M (SD)	range
GPA	1.54 (0.35)	1.0-2.4	1.97 (0.58)	1.0-3.5	2.41 (0.52)	1.3-3.6
math grade	13.75 (1.30)	9-15	12.42 (2.47)	4-15	9.95 (2.43)	4-15
timss	2.98 (1.00)	0-5	2.57 (1.18)	0-5	1.73 (1.07)	0-4
	%		%		%	
school type1	92.98		84.21		63.40	
first1	64.04		63.16		83.01	
tea	53.51		35.53		49.02	

Note. GPA = grade point average in the final exam in school; school type = general-education Gymnasium or not; math grade = math grade in the final exam in school; timss = sum score of the TIMSS items; age = participant's age; first = first mathematical semester or not; tea = teacher candidate or not

For different applications I partition the feature space using a conditional inference tree and assign terminal nodes to the risk group. Results of the decision tree are shown in figure 3.3.

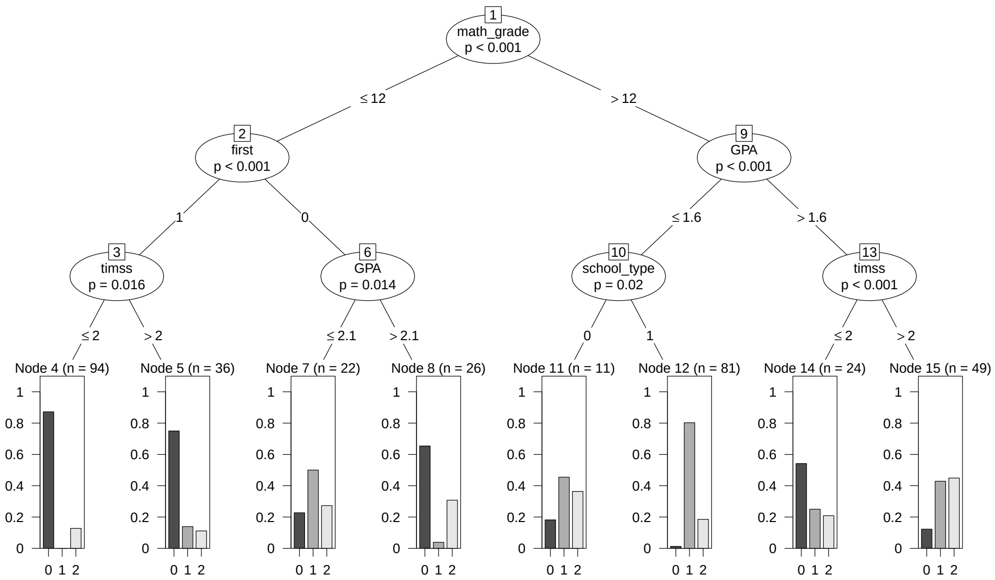

Figure 3.3: Conditional inference tree for the group membership of correct positive prediction, correct negative prediction and wrong predictions an the complete dataset dataC.

Note. The figure shows the split-points within the features used for splits. In the terminal nodes 4, 5, 7, 8, 11, 12, 14, and 15 the frequencies of the group membership is shown, with group 0 being the true negative predictions, group 1 the true positive predictions and group 2 the wrong predictions. The risk group is defined as the true negative predictions. the terminal nodes 4, 5, 8 and 14 are assigned to this group. Predicted outcomes of the ensemble predictor for the complete dataset dataC are used.

GPA = grade point average in the final exam in school; school type = general-education Gymnasium or not; timss = sum score of the TIMSS items; math_grade = math grade in the final exam in school; first = first mathematical semester or not

The figure shows eight terminal nodes. i assign nodes 4, 5, 8 and 14 to the risk groups, because in this nodes the frequency of true negative predictions is outstandingly high compared to the rest. The ensemble predictor is especially confident about the failure of the students in node 4 and node 5. In the following I describe the paths that lead to those terminal nodes. The first split is done at a math grade under and above of 12 points. For students in their first mathematical semester, with a math grade in the final exam in school below or equal to 12 points, the ensemble is confident about the failure in Analysis 1 (regardless of the test performance on the TIMSS items). This path leads to the terminal nodes 4 and 5. For students not in their first mathematical semester a further split is executed at a GPA of 2.1, with GPAs worse than 2.1 leading to terminal node 8, which also is assigned to the risk group.

Students with math grades better than 12 points, are also allocated by their GPA. For students with a GPA worse than 1.6 the math performance is re-assessed with the TIMSS items. Students with 2 correct answers or less are assigned to terminal node 14 and thus to the risk group. Another interesting result shown in figure 3.3 is the relevance of the school type. According to the results, the success in the Analysis 1 lecture for students with good school grades (more than 12 points in the final math exam and a GPA better than 1.6) highly depends on the school type. If the grades are achieved at a general-education Gymnasium, the prediction of success is very confident. For other school types the frequency of positive predictions is still the highest, but by far not as confident. This shows that in this analysis the value of school grades highly depends on the school type.

In summary there are three paths that leading to the risk group, which suggest three especially unfortunate prerequisite constellations. We will summarize these constellations in three risk levels, with the highest risk in risk level one. Students assigned to risk level one have 12 points or below in the final math exam and a GPA higher than 2.1. Risk level two is defined by math grades of 12 points or below and a GPA better than 2.1. Those students might have problems completing the Analysis 1 lectures in the first semester. The risk level three contains students with good math grades (above 12 points) but GPAs not better than 1.6. As already mentioned, the math performance is re-assessed with TIMSS items for this group – 2 or less correctly answered items lead to as risk. The risk levels are summarized in table 3.30.

Table 3.30: Description of three risk levels with especially unfortunate prerequisites.

	risk level one	risk level two	risk level three
math grade	≤12 points	≤12 points	> 12 points
GPA[1]	>2.1	≤ 2.1	> 1.6
timss			≤ 2
description	highest risk level with the most confident failure prediction	high risk level if the Analysis 1 lecture is attended in the first mathematical semester	moderate risk level

Note. [1] the range for the GPA is 1-6 with 1 indicating the best performance
GPA = grade point average in the final exam in school; math grade = math grade in the final exam in school; timss = sum score of the TIMSS items

3.4 Discussion

With regard to the three research questions the discussion is structured as follows. First, I review the results of students* different prerequisites prior to university. I primarily focus on the teacher candidates and their possible differences to the other math students. Secondly, I summarize the most predictive features found by the different methods and thirdly, I discuss the prediction accuracies achieved only using our small set of variables and the risk groups.

3.4.1 Differences in prerequisites of teacher candidates

I compared the prerequisites of B.Ed. students with those of B.Sc. students. Because of the different distribution of students' sex in those groups, with more female students in the B.Ed. program (about 58%) than in the B.Sc. program (about 28%), I repeated the analysis for the performance measures 'GPA', 'math grade' and 'timss' and controlled for sex. The results for the group differences did not change, so those results are not reported. For the GPA, I can reproduce the result of Klusmann et al. (2009), showing no differences for the two groups. Thus a negative selection concerning the general performance in school is not existent in the data. Other than their results, there are differences in the math specific performance measures 'math grade' and ‘timss’, where the B.Sc. students performed better. This might be due to different sample differences used in this study and in Klusmann et al. (2009). I compare teacher candidates in math with B.Sc. students in math and physics, whereas Klusmann et al. (2009) uses the TOSCA data set (Köller et al., 2004), where B.Sc. students with scientific subjects, also outside of the math and physics field are included as well. This means that the different results might occur because of a substantial positive selection in my B.Sc. group. The group differences regarding participation in a mathematical prep course prior to university, with only 20% of the B.Ed. students compared to 50% of the B.Sc. students, need further investigation. The University of Tübingen offers mathematical prep courses for physics and computer science, but no course particular for math students. Even though those courses are open to math students as well, they might be more present in the physics study recommendations, thus the physics students within the B.Sc. group might cause this difference. In the framework of this study the difference of the variable ‘first’ needs to be discussed. Only 58% of the B.Ed. students were in their first mathematical semester, compared to 86% of the B.Sc. students. The data doesn’t contain the information, if the students who are not in their first mathematical semester, already participated in an Analysis 1 lecture before, or if they for example attended lectures in linear algebra first. Due to our exclusions we can eliminate partly successful participations in former Analysis 1 lectures. It is a possible scenario for B.Ed. students not to start with the analysis and linear algebra lectures simultaneously, which would be the common way especially for the B.Sc. students. This has an influence on the dropouts in this data as seen in the importance of the feature ‘first’ in the predictions and feature selections. This might be due to general experience gained at university, even though the lecture itself is attended for the first time. For that reason it was important to include the feature ‘first’ to capture this effect when evaluating the group of teacher candidates.

3.4.2 Identification of the most predictive features

For the identification of the most predictive features method specific algorithms are used. Even though different feature subsets are selected (for example due to the amount of regularization being used in the logistic regression with elastic net regularization), the best predictors show the same subset. The methods of logistic regression (significant coefficients), the linear SVM (recursive feature elimination – RFE), the random forest (conditional variable importance) as well as the two overall best predictors, the logistic regression with elastic net regularization ($\alpha = 0.3$ – selection due to the property of the included L_1 regularization) and the SVM with RBF kernel (generalized RFE), use the features grade point average of the final exam in school ('GPA'), the information if the school type where the final exam is done was a general-education Gymnasium ('school type'), the math grade in the final exam in school ('math grade'), the sum score of the TIMSS items ('timss'), the information if the students were in their first mathematical semester ('first') and if the students are in the B.Ed. (teacher candidates) or in the B.Sc. study program ('tea').

As expected the math specific performance measures and the general school performance measure show positive effects on the success in the Analysis 1 lecture (see e.g. Bean, 2005). As mentioned earlier there is a positive effect concerning students' number of mathematical semester. Students which are not in their first mathematical semester show higher success probabilities than students in their first semester. This effect was expected because even though the students might not have participated in a former Analysis 1 lecture they don't have to deal with general problems at the beginning of university and might be more experienced in handling the requirements of a math lecture. Even though in the basic logistic regression being a teacher candidate shows only just no significant coefficient for the dependent variable indicating the success in Analysis 1, in prediction, all methods selected 'tea' as a predictive feature. Over all, we see slight indication of a positive effect of 'tea' on the success. For example, the single decision tree selects 'tea' as one of the splitting features with a positive effect. This might be due to the high frequency of students not in their first semester within the teacher candidates. Since the feature 'first' is not selected by the tree, its positive effect might occur within the feature 'tea'.

Students who graduated at a general-education Gymnasium are more likely to pass the Analysis 1 lecture. This influence of the school type can only be discussed speculative because the data doesn't reveal further information. In some of the alternative school types the students attend a lower level school first and then transfer to a school were the admission to a university can be achieved. Those differences, for example in the math knowledge, might not be seen in the restricted framework of the final exams in school, but seems to have an influence on success probabilities at the university.

3.4.3 The gained prediction accuracy

Except for the basic logistic regression and the single conditional inference tree, the best predictors of the respective methods all achieve accuracies on the test set above 70%. The F1-scores are in a good range between 0.6 and 0.73 for all predictors (except for the single tree) and show no substantial tradeoff between recall and precision, indicating, as expected, no serious effect of the slightly skewed data.

The overall best predictors are the elastic net – logistic regression with the elastic net regularization and a tradeoff parameter between L_1 and L_2 regularization of $\alpha = 0.3$ – (test accuracy of 74%), the SVM with RBF kernel (test accuracy of 75%) and the ensemble predictor which consists of all the best predictors (test accuracy of 75%). All predictors achieve lower accuracies on cohort 1, but the named best predictors are with accuracies of 71%, for both the SVM and the elastic net, still in a good range. Note that those accuracies are achieved by only using the most predictive features reported in the previous section. The inter-rater agreement values (Cohen's κ) between the predicted outcomes on the test set and the true outcomes are with 0.47 for the elastic net and 0.50 for the SVM in the moderate range (see Landis & Koch, 1977).

As a result I summarize that with appropriate methods the success in the Analysis 1 lecture can be correctly predicted for 75% of the students, only with the knowledge of their GPA, their math grade in the final exam in school, the test result of the TIMSS items, the school type and their number of semesters and study program. Note that this is the accuracy on the test set meaning after possible generalization errors.

3.4.4 Risk groups

The results of this study can help universities to identify risk groups in math study programs. Here we see that unfortunate performance measures from school lead to the expected risk of a failure in Analysis 1. The results not underline the results of previous studies (see e.g. Bean, 2005), but furthermore give thresholds for the math grade of 12 points and a school GPA of 2.1. We also see that a good mathematical performance in school ($>$ 12 points) needs to be confirmed by additional mathematical performance tests, like the TIMSS items (at least as long as the GPA is not in the excellent range, here $>$ 1.6).

3.4.5 Conclusion

I conclude, that teacher candidates start with adverse prerequisites concerning math specific performance measures. However, there are no significant differences to B.Sc. students in terms of success in the Analysis 1 lecture. Success in the Analysis 1 depends on the number of the mathematical semesters with a positive effect of not being in the first semester. The distribution of this variable within the teacher candidates might contribute to a slight overall positive effect of being a teacher candidate. The analysis of the risk groups as well indicates the disagreement with the public opinion of teacher candidates being the worse students, described in Blömeke (2005). The study program itself does not occur as indicator for a risk group. But one should mention that a threshold for the math grade in school can be set at 12 points (see previous section). The B.Ed. students however show significantly worse math grades compared to their B.Sc. colleagues (11.43 and 12.06 respectively) with the mean lying slightly under the threshold. An assumable negative effect on the success however did not occur in our analysis and even though the teacher candidates show worse prerequisites in math performance the school grades are still in the good range. As a conclusion I say that at least in our dataset the teacher candidates show no substantially unfortunate prerequisites for the success in the Analysis 1 lecture.

With regard to the variable clusters of Bean (2005) and Burrus et al. (2013) the analysis showed not effects in the field of students demographic characteristics, age and sex, but the a high relevance of the academic preparation and success factors. The indication of risk groups highly depends on performance measures from school, or at least on their underlying concepts of knowledge. We saw that the school grades in some cases have to be confirmed by additional tests or the knowledge about the school types where the grades were achieved. The identification of the risk group indicators and the classification of the risk levels can be extremely helpful for universities in the discussion about admission restrictions. Table 3.30 shows the highest risk for students with below average school grades (math grades of ≤ 12 points and GPAs higher than 2.1). Especially the math grade seems to be a good indicator for students in risk groups. Other applications than admission restrictions can be the development of interventions and general support courses. Even though, the results might not give suggestions for specific variables which could be improved by interventions, it helps to identify risk groups for which interventions or support courses should be developed. For this task figure 3.3 can be used.

This analysis relies on results of the ensemble predictor, which shows a test accuracy of 75%. This accuracy is only achieved using a small set of features, consisting of information prior to the lecture. Considering no information about students' behavior during the semester is included, this is a rather high value. Note that the accuracy refers to predictions on the test set, consisting of students the algorithm has never seen bevor. The remaining 25% of the students show wide ranges of variable values. Again, this stresses the importance of the behavior during the semester. On one hand, even students with very unfortunate prerequisites can succeed in the lecture and, on the other hand, very fortunate prerequisites don't guarantee the success. With the high accuracy of 75% a general structure seems to be found, resulting, for example, in the identification of risk groups.

3.5 Limitations and outlook

In this study, data of two consecutive Analysis 1 lectures at the University of Tübingen was used. Even though, instead of sampling, all students of the lecture were included (with a return rate around 92%), the data might only be representative for this university. Especially concerning the federal state, where 87% of the participants graduated at a school in Baden Württemberg, the data might not be representative for Germany. With the combination of two lectures I tried to address the possible dependence on lecturers and specific schedules. For further improvement the inclusion of more lectures, in particular lectures at different universities, would be appropriate. Concerning the used variables, one should consider aspects concerning students' personality and motivation.

As next steps the inclusion of behavioral information during the semester is planned, including a wider range of students' characteristics. For a better generalization more universities in Baden Württemberg will be included.

References

Adams, R. J., Wu, M. L., & Macaskill, G. (1997). Scaling methodology and procedures for the mathematics and science scales. *TIMSS technical report*, *2*, 111–145.

Ball, D. L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (Vol. 4, pp. 433 – 456). New York, NY: Macmillan.

Baumert, J., et al. (Eds.). (1999). *Testaufgaben zu TIMSS/III: Mathematisch-naturwissenschaftliche Grundbildung und voruniversitäre Mathematik und Physik der Abschlussklassen der Sekundarstufe II (Population 3)* (Vol. 62). Berlin: Max-Planck-Institut für Bildungsforschung.

Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 29 – 53). Waxmann.

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, *12*(2), 155–187.

Bean, J. P. (1983). The application of a model of turnover in work organizations to the student attrition process. *The Review of Higher Education*, *6*(2), 129–148.

Bean, J. P. (2005). Nine themes of college student retention. In A. Seidman (Ed.), *College student retention: Formula for student success* (p. 215-

244). New York, NY: Rowman & Littlefield.

Blömeke, S. (2005). Das Lehrerbild in Printmedien. *Die Deutsche Schule*, *97*(1), 24–39.

Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees.* Wadsworth, California: Taylor & Francis.

Burrus, J., Elliott, D., Brenneman, M., Markle, R., Carney, L., Moore, G., ... Roberts, R. D. (2013). Putting and keeping students on track: Toward a comprehensive model of college persistence and goal attainment. *ETS Research Report Series*.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 27:1–27:27.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187-220.

Eilerts, K. (2009). *Kompetenzorientierung in der Mathematik-Lehrerausbildung: empirische Untersuchung zu ihrer Implementierung* (Vol. 14). Münster: LIT Verlag.

Evgeniou, T., Pontil, M., & Elisseeff, A. (2004). Leave one out error, stability, and generalization of voting combinations of classifiers. *Machine Learning*, *55*(1), 71–97.

Feldman, M. J. (1993). Factors associated with one-year retention in a community college. *Research in Higher Education*, *34*(4), 503–512.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22.

Grünwald, N., Kossow, A., Sauerbier, G., & Klymchuk, S. (2004). Der Übergang von der Schul- zur Hochschulmathematik: Erfahrungen aus internationaler und deutscher Sicht. *Global Journal of Engineering Education*, *8*(3), 283–293.

Gueudet, G. (2008). Investigating the secondary–tertiary transition. *Educational Studies in Mathematics*, *67*(3), 237–254.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002, Jan 01). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1), 389–422.

Hagedorn, L. S., Maxwell, W., & Hampton, P. (2001). Correlates of retention for african-american males in community colleges. *Journal of College Student Retention: Research, Theory & Practice*, *3*(3), 243-263.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York, NY: Springer.

Heublein, U. (2014). Student dropout from german higher education institutions. *European Journal of Education*, *49*(4), 497-513.

Heublein, U., Hutzsch, C., Schreiber, J., Sommer, D., & Besuch, G. (2009). *Ursachen des Studienabbruchs in Bachelor- und in herkömmlichen Studiengängen: Ergebnisse einer bundesweiten Befragung von Exmatrikulierten des Studienjahres 2007/08.* Hannover: HIS Hochschul-Informations-System.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. (2006). Survival ensembles. *Biostatistics*, *7*(3), 355–373.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674.

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd ed.). New York, NY: Springer.

Klusmann, U., Trautwein, U., Lüdtke, O., Kunter, M., & Baumert, J. (2009). Eingangsvoraussetzungen beim Studienbeginn: Werden die Lehramtskandidaten unterschätzt? *Zeitschrift für Pädagogische Psychologie*, *23*(34), 265–278.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1), 273 - 324. (Relevance)

Köller, O., Watermann, R., Trautwein, U., & Lüdtke, O. (Eds.). (2004).

Wege zur Hochschulreife in Baden Württemberg. TOSCA – Eine Untersuchung an allgemeinallgemein und beruflichen Gymnasien. Opladen: Leske + Budrich.

Krauss, S., Neubrand, M., Blum, W., & Baumert, J. (2008). The professional knowledge of german secondary mathematics teachers: Investigations in the context of the COACTIV project. In *11th International Congress on Mathematics Education (ICME 11).* Monterrey, Mexico.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV.* Waxmann.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

LeCun, Y., Denker, J. S., & Solla, S. A. (1990). Advances in neural information processing systems 2. In D. S. Touretzky (Ed.), (pp. 598–605). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Leppel, K. (2002). Similarities and differences in the college persistence of men and women. *The Review of Higher Education*, *25*(4), 433-450.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2017). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien [Computer software manual]. (R package version 1.6-8)

Mullis, I. V., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A., & Erberber, E. (2007). *TIMSS 2007 assessment frameworks.* Lynch School of Education: Boston College: TIMSS & PIRLS International Study Center.

Nagy, G. (2007). *Berufliche Interessen, kognitive und fachgebundene Kompetenzen: Ihre Bedeutung für die Studienfachwahl und die Bewährung im Studium* (phd thesis, Freie Universität, Berlin). Retrieved from `http://www.diss.fu-berlin.de/2007/109/`

Quinlan, J. R. (1993). *C4.5: Programs for machine learning.* San Francisco, California: Morgan Kaufmann Publishers Inc.

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Rakotomamonjy, A. (2003). Variable selection using svm based criteria. *Journal of Machine Learning Research*, *3*, 1357 - 1370.

Schiefele, U., Streblow, L., & Brinkmann, J. (2007). Aussteigen oder Durchhalten. Was unterscheidet Studienabbrecher von anderen Studierenden? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *39*(3), 127-140.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(307).

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*(25).

Strobl, C., Hothorn, T., & Zeileis, A. (2009). Party on! *The R Journal*, *1*(2), 14–17.

Terenzini, P. T., & Pascarella, E. T. (1980). Toward the validation of tinto's model of college student attrition: A review of recent studies. *Research in Higher Education*, *12*(3), 271–282.

Terry M. Therneau, & Patricia M. Grambsch. (2000). *Modeling survival data: Extending the Cox model*. New York: Springer.

Therneau, T. M. (2015). A package for survival analysis in s [Computer software manual]. (version 2.38)

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), 89-125.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition*. Chicago, IL: University of Chicago Press.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago, IL: University of Chicago Press.

Chapter 4

Discussion

In this thesis three main aspects of math teacher education are covered. The inner structure of MPCK and a potential separation to MCK is investigated in chapter 1. Building on these results, in chapter 2 this relationship is amplified and prerequisites students enter university with for the acquisition of MCK and MPCK are explored. Because of the high relevance of MCK, both for teaching and as requirement for MPCK, in chapter 3 I examined the success, in terms of dropout rates, of students in a first semester mathematical lecture with emphasis on influencing factors and risk groups. An important question of this chapter is, if teacher candidates show disadvantageous prerequisites for the success in the math lecture and thus build a risk group. While chapter 1 and chapter 2 rely directly on the performance on MCK and MPCK tests, chapter 3 does not focus on the competence itself, but on the successful participation in a lecture for the acquisition of those competences, or at least MCK.

The validity of the performance tests was ensured in several steps. Both tests were developed in close collaboration with experts in the specific field to assure content validity. The correlation to validity items employed in TEDS-sM (Buchholtz et al., 2012) and TIMSS (Baumert et al., 1999) was substantial indicating convergent validity. Additionally, correlations to school grades, such as the GPA and the final high school math score, as well as mean differences on the scales met expectations.

Inner facets of MPCK

In chapter 1 I focus on the structure of MCK and MPCK, in particular on the structure within the latter. Model results favored a model that distinguishes two separate facets within MPCK (see figure 1.6). This model supports the existence of the independent MPCK facets *instruction* and *diagnostic competence* in addition to the general MCK facet. Here again the framework is the subject matter dominated, initial phase of the study program. Even without a special education in MPCK, those facets were able to explain variance in a content-related MPCK test, additionally to a general mathematical understanding. The statistical separation of inner MPCK facets has not been done before and thus is remarkable, especially in this content-related context. The results not only show independent inner facets of MPCK, but also the separability of those facets from MCK.

The separation of the inner facets of MPCK is interesting from a theoretical point of view, with respect to the emergence and structure of competencies. This knowledge can be applied in the planning of lectures and seminars.

Relation of MCK and MPCK

As mentioned before, in addition to inner facets of MPCK, the separability of MCK and MPCK is shown in chapter 1. Although the separation of MCK and MPCK has already been analyzed in previous studies (see e.g. Buchholtz, Kaiser, & Stancel-Piatak, 2011), the results of this study are remarkable, because the separation of MPCK and MCK was undertaken with a highly content-related point of view on MPCK. This shows, that there might be a generic separation of MPCK and MCK in the sense of Shulman's PCK as "subject matter knowledge for teaching" (Shulman, 1986, p. 9). As discussed in chapter 1, this goes beyond the separation of MCK and a comprehensive MPCK dimension – including not only content-related parts, but also the general pedagogical point of view – at a later point in the training as employed by the TEDS-group (e.g. Blömeke et al., 2011; Blömeke, Kaiser, & Lehmann, 2010; Buchholtz et al., 2012) and COACTIV (Krauss, Neubrand, Blum, & Baumert, 2008; Kunter et al., 2011). A separation in their broader context seems more obvious, because they included aspects of MPCK, which differ greatly from MCK. To show this generic separation in the content-related context, the study was deliberately scheduled at this early stage of the training. The results point to the mindset of different kinds of mathematical understanding – referred to as MPCK and MCK – at the beginning of the training.

Even though MCK and MPCK can be statistically separated, results show the expected substantial relation in terms of significant correlations (see chapter 2). However, the correlation was low enough for the assumption of two different and separable dimensions. The overlap of the facets can also be seen in the similarity of the effect of covariates as conditional factors for MCK and MPCK.

In addition to MCK, MPCK is important for teachers and according to the results of chapter 1, it starts to develop together with MCK from the beginning of the training. The separation, however, shows that it is not one simple unidimensional construct and thus MPCK should be more stressed in addition to the mathematics lectures (MCK) early in the training (chapter 1).

Prerequisites for the acquisition of MCK and MPCK

Students' prerequisites for the acquisition of MCK and MPCK are examined in chapter 2. Additional statements about the prerequisites for the acquisition of MCK, in terms of conditional factors for the success in a MCK lecture, are discussed in chapter 3.

While a dependence of students' sex on the MCK test results was found (chapter 2), students' sex seems to be unrelated to the success in the mathematical lecture (chapter 3). For both, the MCK test and the success in the mathematical lecture, the school performance measures showed significant coefficients.

In both studies the school type also showed significant coefficients. In contrast to the MCK, the MPCK performance was unrelated to the school type, whereas students' sex, GPA and math grade showed the same dependence (with male students and better school performances being favored).

Better performance in school not only leads to better performance in MCK, but also in MPCK (even though it is not part of school education). This supports the importance of mathematical performance as requirement for MPCK (chapter 2). The effect of the experience of students leads in the same direction. Mathematically more experienced students did not only show better results on the MCK test, but also on the MPCK test, even though no particular MPCK lectures or seminars were attended. Once more, this might be a sign of the role of MCK as a prerequisite for MPCK.

For both tests the study program showed no significant effect, meaning no differences between teacher candidates and their colleagues at the beginning of the training were detected. This is interesting, because the choice of the study program Bachelor of Education, and a thereby probably involved affinity to teach, does not seem to have an effect on the performance on the MPCK test.

According to group differences prior to university (see chapter 3), teacher candidates show no difference in their GPAs compared to their colleagues. This does not apply for the mathematical performance measures, math grade in school and the results of the TIMSS items. Those measures show significant differences between B.Ed. and B.Sc. students, with B.Sc. students showing the better performance. Note, that the TIMSS items were part of the MCK test. TIMSS items form the part of the MCK test, which measures mathematical knowledge from school, whereas the remaining part of the MCK test measures mathematical abilities related to upcoming contents at university. Differences were only seen concerning mathematical school performances.

Acquisition of MCK in a first semester mathematical lecture

In chapter 3 I analyze the success in a first semester mathematical lecture, depending on numerous prerequisites. I use two cohorts of the Analysis 1 lecture, which is the typical entry lecture for both B.Ed. and B.Sc. students at the University of Tübingen. For the dropout prediction students' GPA, math grade, performances on the TIMSS items as well as the study program, the school type and the variable 'first' – indicating whether the students are in their first mathematical semester or not – were the most predictive features. Note that the importance of the GPA, school type and the math grade are in accordance with the results of chapter 2 addressing prerequisites for the acquisition of MCK. All of the used features represent student attributes prior to the lecture. A prediction accuracy of 75% was reached, which is remarkable no information about students' behavior during the semester was included. This illustrates the dependence of success in this lecture on prerequisites and allows definite statements on the identification of risk groups at the beginning of studies already. The analysis of those risk groups provides interesting insights, for university scholars as well as the department of education. The analysis does not only rely on linear models like the logistic regression, but includes more complex relations of the independent variables with success and between the variables. Students with a math grade lower or equal to 12 points in the final exam in high school and a GPA higher (worse) than 2.1 show the highest risk level with the most confident failure prediction. This again shows the dependence of school performance measures not only on MCK (see chapter 2), but also on the success in the lecture. This expected dependence was mentioned before (see e.g. Bean, 2005) but without precise thresholds.

Another interesting result concerns students with a good mathematical performance in school (math grade). In order to be predictive for the success, the mathematical school performance needs further confirmation with the TIMSS items (at least as long as the GPA is not in the excellent range ≤ 1.6). This result questions the validity of the math grade as a measure for mathematical knowledge. Additionally the predictive character of school performance measures depends on the school type, meaning that the explanatory power of math grade and GPA vary between school types.

The selection of the study program as a predictive feature needs further investigation. The effect of the study program B.Ed. is rather positive. This might be due to the high frequency of students not in their first semester within the teacher candidates. The analysis of the risk groups, survival models and the basic logistic regression did not reveal an effect of the study program.

Conclusion

The importance of both MCK and MPCK for teaching math has been reported several times before (see e.g. Ball, Lubienski, & Mewborn, 2001; Shulman, 1987). In this thesis the dependence of those facets of knowledge was confirmed. Despite the dependence of the facets, I also showed their distinguishability in a highly subject matter dominated phase of teacher eduction. Those findings recommend the promotion of MPCK already in an early phase of studies and parallel or supporting to the MCK lectures. MPCK as another kind of mathematical understanding is a content-related mindset that is often not considered.

A broad differentiation of contents within MPCK has been suggested before (see e.g. Blömeke & Kaiser, 2014; Döhrmann, Kaiser, & Blömeke, 2012; Krauss et al., 2011; Krauss, Brunner, et al., 2008; Tatto et al., 2008). I summarized those suggestions with a content-related point of view and introduced two knowledge facets within MPCK (*instruction* and *diagnostic competence*). I was able to verify those facets empirically and their identification in addition to the general MCK (chapter 1).

On MCK, I saw advantages of the B.Sc. students, compared to the B.Ed. students, in terms of mathematical school performance prerequisites (chapter 3). The performance on (only) the TIMSS items showed the same results, whereas the performance on the complete MCK test (as well as the GPA) revealed no difference between the two groups (chapter 2). I conclude that there are differences in the previous mathematical school knowledge, which do not seem to persist for a mathematical knowledge on a higher abstraction level, but still based on school contents. This result was confirmed in the analysis of dropout rates in chapter 3.

In summary the teacher candidates do not form a risk group in math education.

References

Ball, D. L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (Vol. 4, pp. 433 – 456). New York, NY: Macmillan.

Baumert, J., et al. (Eds.). (1999). *Testaufgaben zu TIMSS/III: Mathematisch-naturwissenschaftliche Grundbildung und voruniversitäre Mathematik und Physik der Abschlussklassen der Sekundarstufe II (Population 3)* (Vol. 62). Berlin: Max-Planck-Institut für Bildungsforschung.

Bean, J. P. (2005). Nine themes of college student retention. In A. Seidman (Ed.), *College student retention: Formula for student success* (p. 215-244). New York, NY: Rowman & Littlefield.

Blömeke, S., et al. (Eds.). (2011). *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT.* Münster: Waxmann.

Blömeke, S., & Kaiser, G. (2014). Theoretical framework, study design and main results of TEDS-M. In S. Blömeke, F.-J. Hsieh, G. Kaiser, & W. H. Schmidt (Eds.), *International perspectives on teacher knowledge, beliefs and opportunities to learn: TEDS-M results.* Springer.

Blömeke, S., Kaiser, G., & Lehmann, R. (Eds.). (2010). *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich.* Münster: Waxmann.

Buchholtz, N., Kaiser, G., & Stancel-Piatak, A. (2011). Professionelles Wissen von Studierenden des Lehramts Mahtematik. In S. Blömeke et al.

(Eds.), *Kompetenzen von Lehramtsstudierendenin gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT* (pp. 101–133). Waxmann.

Buchholtz, N., Scheiner, T., Döhrmann, M., Suhl, U., Kaiser, G., & Blömeke, S. (2012). *TEDS-shortM: Kurzfassung der mathematischen und mathematikdidaktischen Testinstrumente aus TEDS-M, TEDS-LT und TEDS-Telekom.* Hamburg: Universität.

Döhrmann, M., Kaiser, G., & Blömeke, S. (2012). The conceptualisation of mathematics competencies in the international teacher education study TEDS-M. *ZDM*, *44*(3), 325–340.

Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., ... Elsner, J. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 135 – 162). Waxmann.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, *100*(3), 716 – 725.

Krauss, S., Neubrand, M., Blum, W., & Baumert, J. (2008). The professional knowledge of german secondary mathematics teachers: Investigations in the context of the COACTIV project. In *11th International Congress on Mathematics Education (ICME 11).* Monterrey, Mexico.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Waxmann.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review*, *57*(1), 1–23.

Tatto, M. T., Ingvarson, L., Schwille, J., Peck, R., Senk, S. L., & Rowley, G. (2008). *Teacher education and development study in mathematics*

(TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. conceptual framework. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.

Appendix A

Example tasks for the performance tests

A.1 Example Tasks for the pedagogical content knowledge

Example tasks for the MPCK performance test used in chapter 1 and chapter 2. On example for each of the three facets *school-relevant mathematical content knowledge* (schoolMCK), *diagnostic competence* and *instruction competence*. The test was applied in German, thus we provide the original version in German and a translated version in English.

A.1.1 School-relevant mathematical content knowledge

Original version in German:

In einer Klausur wird Schülern (für feste $a, b \in \mathbb{R}$) der Ausdruck

$$\int_a^b f(x)dx$$

gezeigt. Die Frage ist, was dieser aussagt?

Entscheiden Sie welche Antwort korrekt ist.

Kreuzen Sie ein Kästchen pro Zeile an

	Der Ausdruck steht...	Ja	Nein
A)	...immer für eine Stammfunktion von f, die an einer Stelle ausgewertet wurde.	☐	☐
B)	...für den Flächeninhalt zwischen dem Graphen von f und der x-Achse, falls $f(x) > 0, \forall x \in [a, b]$.	☐	☐
C)	...für eine Funktion F, für die gilt: $F' = f$.	☐	☐

English version:

In an exam the term

$$\int_a^b f(x)dx$$

is shown to the students (for fixed $a, b \in \mathbb{R}$. The question is what it means. Decide which answer is correct

Mark one box per row

	The term indicates...	Yes	No
A)	...always an antiderivative of f evaluated at one point.	☐	☐
B)	...the area between the graph of f and the x-axis, if $f(x) > 0, \forall x \in [a, b]$.	☐	☐
C)	...a function F with $F' = f$.	☐	☐

A.1.2 Diagnostic competence facet

Original version in German:

Ein Schüler soll das Fassungsvermögen (in Litern) des Tanks eines Autos berechnen. In der Aufgabe ist ein Verbrauch von 7,6 Litern pro 100 km und eine maximale Reichweite von 530 km angegeben.
In der 7. Klasse hat Peter die folgende, falsche Antwort gegeben:

Nach 100 *km hat das Auto* 7,6 *Liter verbraucht. Damit kann das Auto mit einem Liter* 13,16 *km weit fahren. Für die angegebenen* 530 *km muss der Tank daher* $13,16 \cdot 530 = 6974,8$ *Liter fassen.*

Bei welchen der folgenden Aufgaben A,B und C besteht die Möglichkeit, dass Peters Fehler wiederum zu einer falschen Lösung führt?

Kreuzen Sie ein Kästchen pro Zeile an

		Ja	Nein
A)	Kürzen Sie den Bruch $\frac{3x^2}{2x+8x^3}$ vollständig.	☐	☐
B)	Wie viel sind 13 % von 120€?	☐	☐
C)	Geben Sie alle $x \in \mathbb{R}$ an, die die Gleichung $5,6x - 12 = 0$ lösen.	☐	☐

English version:

A student is asked to calculate the capacity of a car's tank (in liter). A fuel consumption of 7.6 liters per 100 km and a maximum range of 530 km are given in the task.
In seventh grade Peter gave the following, wrong answer:

After 100 *km the car has used* 7.6 *liters. Therefor the car can go* 13.16 *km on one liter. Thus for the* 530 *km given in the task, the capacity of the tank hast to be* $13.16 \cdot 530 = 6974.8$ *liters.*

For which ones of the following tasks is it possible that Peter's mistake would also lead to a wrong answer?

Mark one bos per row

		Yes	No
A)	Reduce the fraction $\frac{3x^2}{2x+8x^3}$ completely.	☐	☐
B)	How much is 13 % of 120€?	☐	☐
C)	Give all $x \in \mathbb{R}$ which solve the equation $5,6x - 12 = 0$.	☐	☐

A.1.3 Instruction facet

Original version in German:

Sie werden in der Schule darauf angesprochen, warum Sie immer die reellen Zahlen $\mathbb{R}$ verwenden, wenn irrationale Zahlen meist keine direkte Rolle spielen und deshalb die rationalen Zahlen $\mathbb{Q}$ doch reichen würden.
Welche Erklärungen wären angemessen?

Kreuzen Sie ein Kästchen pro Zeile an

		Ja	Nein
A)	Das macht rein pädagogisch Sinn! Es wurde nachgewiesen, dass es Schülern leichter fällt bestimmte Rechnungen in der Dezimalschreibweise durchzuführen, im Gegensatz zu komplizierten Bruchdarstellungen.	☐	☐
B)	In der Schule werden oft Längen, Flächeninhalte und Volumina gemessen. Das setzt einen Zahlenbereich voraus, der in eineindeutiger Beziehung zu den Punkten auf einer Geraden steht. Das leisten die reellen Zahlen im Gegensatz zu den rationalen Zahlen.	☐	☐

English version:

The question arises why real numbers $\mathbb{R}$ are always used in school even though irrational numbers hardly ever occur, which means that using rational numbers $\mathbb{Q}$ should be enough.
Which explanation would be appropriate?

Mark <u>one</u> box <u>per row</u>

		Yes	No
A)	This makes sense from a pedagogical point of view! It has been shown that certain calculations are easier for students in decimal notation compared to complicated fractions.	☐	☐
C)	In school lengths, areas and volumes are frequently measured. This requires a number range which is in a biunique relation to the points on a line. This is provided by real numbers rather than by rational numbers..	☐	☐

A.2 Example task for the content knowledge

Example task for the MCK performance test used in chapter 2. Additionally to the applied German version we provide a version in English.

Original version in German:

Funktionen sind Ihnen zum Beispiel in der Form $f(x) = x^2$ bekannt. Welche Vorstellung haben Sie von einer Funktion?

Kreuzen Sie <u>ein</u> Kästchen <u>pro Zeile</u> an

		wahr	falsch
A)	Eine Funktion f ist eine Zuordnung zwischen zwei Mengen A und B, die jedem Element a aus A genau ein Element b aus B zuordnet.	☐	☐
B)	Eine Funktion ist ein formaler Ausdruck in einer Unbestimmten x, deren Definitionsbereich noch bestimmt werden muss.	☐	☐
C)	Haben zwei Funktionen den gleichen Graphen, so sind sie gleich.	☐	☐

English version:

You know functions for example in the form $f(x) = x^2$. Which conception of functions do you have?

Mark <u>one</u> box <u>per row</u>

		true	wrong
A)	A function f refers to an assignment between two sets A and B, which assigns to each element a in A exactly one element b in B.	☐	☐
B)	A function is a formal expression of one variable x which domain has to be identified.	☐	☐
C)	Two functions with the same graph are equal.	☐	☐